Productivity Beliefs and Efficiency in Science

Fabio Bertolotti* Kyle Myers† Wei Yang Tham‡

November 13, 2024

Abstract

We develop an approach for estimating producers' productivity beliefs without observing their output, and we use it to evaluate allocative efficiency in science. We build a model of researchers' utility functions, which implies that researchers' willingness to pay for the two inputs of the scientific production function—funding and time—reveals their productivity beliefs. We estimate the parameters of the model using rich data from a nationally representative survey of research-active professors from all major fields of science. The research productivity distribution is very skewed even within narrowly defined fields. Using the productivity estimates, we provide a systematic view of efficiency in science by comparing the actual allocation of inputs to the optimal allocations given researchers' beliefs and some assumed objectives. The market of science appears to be moderately efficient; there is a significant positive correlation between researchers' actual input levels and their optimal input levels implied by the model. However, the wedge between researchers' actual and optimal input levels is often significant, and it is difficult to use conventional observable measures to predict which researchers are over- or under-resourced. We also explore the degree to which gender biases or an over-reliance on proxies may contribute to misallocation. Valuing the gains from a more efficient allocation is challenging, but our estimates imply that total budgets would need to increase by roughly 40% ($155,000 per researcher per year) under actual allocations in order to achieve the same growth in scientific output that we predict under alternative allocations of the current budget. There are substantial gains from developing new ways to identify productive scientists and ensuring they have access to more funding and time for their work.

*Bertolotti: Bank of Italy, fabiobertolotti@hotmail.it.

†Myers: Harvard Business School & Laboratory for Innovation Science at Harvard, kmyers@hbs.edu.

‡Tham: University of Toronto, weiyang.tham@rotman.utoronto.ca.

This project received financial support from The Alfred P. Sloan Foundation and the Harvard Business School. This project would not have been possible without the excellent work of Rachel Mural, Nina Cohodes, and Yilun Xu as well as input from Marie Thursby, Jerry Thursby, and Karim Lakhani. We have received helpful comments from Judy Chevalier, Daniel Gross, Bronwyn Hall, Petra Moser, David Rivers, as well as seminar participants at JOMT, Massachusetts Institute of Technology, the Workshop on the Organisation, Economics and Policy of Scientific Research, the Banff Productivity, Innovation, and Economic Growth Conference, and the Conference on the Economics of Innovation in Memory of Zvi Griliches.

The science and engineering community tends to feel, with some reason, that every field and sub-discipline is underfunded... but money is not likely to become available to "fully fund" every field even if we could agree on what that would mean.

U.S. Congressman Sherwood Boehlert, 2001
Chairman of the Science Committee

1 Introduction

The aggregate output of any market depends on input levels, producers' productivity, and the extent to which more productive producers have greater access to inputs—the market's allocative efficiency. The advent of large-scale data sets detailing firm-level inputs and outputs, combined with advances in theory and estimation techniques, now allow economists to estimate the distribution of firms' productivity and evaluate how efficiently inputs are allocated across industrial markets. Overall, these analyses indicate that a substantial amount of the differences in output across commercial sectors and geographic regions can be attributed to differences in allocative efficiency.[1]

Virtually no analysis of this sort has been conducted for the market of science.[2] A long line of sociological work has convincingly shown that frictions in science exist to a non-zero degree (e.g., Merton 1973; Zuckerman 1988; Shapin 1995), but we still have a very limited quantitative sense as to how much these frictions are slowing down scientific progress. And a growing body of research has documented significant aggregate returns to investment in science on a per-dollar basis (e.g., Azoulay et al. 2019b), but economists have not had much to say the distribution of productivity or allocative efficiency across science. This is unfortunate given the large role that the institutions of basic science (e.g., universities, national laboratories) have been shown to play in driving commercial R&D and economic growth more generally.[3]

Understanding productivity and efficiency in science has become increasingly important as aggregate R&D productivity continues to decline (Bloom et al. 2020) and the division of labor between academic and commercial science continues to grow (Arora et al. 2020). However, a proper analysis of productivity and efficiency in science using prevailing methodologies (e.g., Olley and Pakes 1996; Hsieh and Klenow 2009) would be exceptionally difficult. In this paper, we develop an alternative approach to estimating productivity and efficiency that sidesteps some challenges while introducing some new ones, and we apply it to the market of academic science. As a result, we can estimate the distribution of scientific productivity, which appears to be very skewed, and quantify the value from more efficiently allocating inputs to the most productive researchers, which appears to be very large.

[1] For example, see: Hsieh and Klenow (2009); Syverson (2011); Restuccia and Rogerson (2013); Hopenhayn (2014).
[2] The closest examples we are aware include Adams and Griliches (1998); Wang and Huang (2007); Rosenbloom et al. (2015), and Qiu (2023), but these evaluations of efficiency are quite indirect.
[3] For example, see: Jaffe (1989); Adams (1990); Mansfield (1998); Jaffe and Lerner (2001); Popp (2017); Valero and Van Reenen (2019); Hausman (2022); Andrews (2023)

At the core of our approach is the following concept: if you know a producer's demand for their factors of production as well as consumers' demand for the output, then you can infer the producer's belief about their productivity without ever observing their output. This concept is useful here because the outputs of science (e.g., new ideas) are notoriously hard to observe, and the best proxies (e.g., publications) are rife with limitations. As Adams and Griliches (1998) put it: "*what constitutes a scientific paper makes for an elastic yardstick of scientific achievement.*"

The first key ingredient to our approach is a new model of researchers' production and consumption decisions. Researchers derive utility from their salary, the output of their scientific production, and their leisure time. They take the following as given: their salary, their research funding guarantees, and their time constraints from required duties (e.g., administration, teaching). Researchers' choice is how to allocate their time across fundraising (e.g., writing grants), research, and leisure.[4] Importantly, each researcher has a unique scientific production function with constant returns to scale that is defined by two researcher-specific parameters: a funding-intensity parameter and a factor-neutral productivity parameter that describes researchers' total factor productivity (TFP).

Using the model, we can write researchers' demand for their two inputs—money and time—as a function of their beliefs about their own productivity. Furthermore, researchers' input demand can be expressed by their willingness to sacrifice part of their salary for the inputs. This tradeoff mirrors real decisions researchers make in the job market, and it forms the basis of the experiments that provide the variation necessary for identifying the model's parameters. There are a number of important limitations to the model that we enumerate below, but it captures many realities of academic science.

While we avoid the need to observe outputs, we still need researchers' salaries, their time allocations, their access to inputs, and their willingness to pay for these inputs. Thus, the second key ingredient to our approach is a novel survey of researchers that contains all of these components. The survey covers a nationally representative sample of research professors across all major fields of science at roughly 150 major institutions of higher education in the US; for more, see Myers et al. (2023).[5] Beyond the necessary observational data, the survey includes a series of hypothetical experiments that solicit researchers' willingness to trade off their salary for more research funding or fewer administrative duties.

Researchers' willingness to pay for inputs are of plausible magnitudes and behavior. The median researcher is willing to pay 10 cents for $1 of research funding and $68 for one less hour of administrative duties. The components of the model explain a large fraction of the variation in researchers' responses (82–96%) and, for the most part, responses do not appear to reflect sample selection

[4] To capture heterogeneity in researchers' preferences and the demand for their output, we allow some functional forms to depend on individuals' observable features.

[5] Evidence is provided in Myers et al. (2023) suggesting the presence of non-response bias in the sample is very low on observable dimensions such as institutional rank and researcher-level variables such as grant funding and publication rates; some of this is reprinted in the Appendix.

or systemic noise in respondents willingness to pay for any hypothetical good.[6] Furthermore, researchers' willingness to pay for free time is strongly correlated with their implied hourly wages as traditional theory would predict.

To estimate the model's parameters, we use Generalized Method of Moments (GMM) to minimize the distance between researchers' answers to the thought experiments and what we'd expect given the observational data, the optimality conditions implied by the model, and a parameter guess. This approach comes with two important caveats. First, while the thought experiments are structured to mimic real-world examples of job offer comparisons, we cannot escape the fact that there are no strong incentives for truthful responses. Our approach allows us to bypass the issue of output measurement, though its validity depends on the quality of our survey responses. Because of this, we pay considerable attention to exploring the plausibility of the magnitudes of our estimates and correlations between our estimates and other real-world metrics from auxiliary data (e.g., publication records). Second, our approach does not identify realized productivity, but instead identifies researchers' *beliefs* about their future productivity. Science is uncertain, so there will certainly be discrepancies between these beliefs and researchers' actual productivity. Furthermore, behavioral biases that can lead certain individuals to over- or under-estimate their own capabilities will generate further discrepancies (e.g., Sarsons and Xu 2021). Still, allocating resources in science always involves a reliance on researchers' beliefs to some degree, since each new research project requires predictions of future productivity. Thus, our focus on productivity beliefs seems warranted.

Our estimates of productivity beliefs vary widely across researchers, even after accounting for outliers. Within major fields of study, the ratio of the 90^{th} and 10^{th} percentile of productivity is approximately 11.6. When we look in narrower fields of study and attempt to control for other sources of heterogeneity, we still estimate 90–10 ratios to be approximately 5.6. We are cautious in comparing our researcher-level productivity metrics to the firm-level productivity metrics others have estimated for industrial sectors, but loosely speaking, this is a very high degree of dispersion compared to what is observed in commercial markets (c.f., Syverson 2011).[7] This degree of skew suggests that it may be hard for the market of science to facilitate positive selection on productivity (e.g., Azoulay et al. 2019a). Thus, we turn next to our analyses of the market's allocative efficiency.

To evaluate allocative efficiency, we compare inputs and outputs under actual allocations to those under alternative allocations. For comparison, we study allocations that achieve three alternative

[6] The survey recruitment included randomized incentives and reminders, which we use to test for sample selection per Heckman (1979), and we also implement Dizon-Ross and Jayachandran's (2022) approach of using a "benchmark good" as a part of our willingness-to-pay elicitation to test whether respondents exhibit systematic noise in their stated preferences.

[7] Some of this variation may also reflect the fact that we are limited in how we allow other parameters in researchers' utility functions to be individual specific. However, a large vector of covariates describing researchers' professional and socio-demographic features explains very little of the variation in their willingness to pay responses beyond the variables of the model, which suggests that un-modeled heterogeneity in our parameters may not be so large. As a signal of face validity, our productivity estimates are positively correlated with common metrics of scientific productivity (e.g., publications, citations, grant funding).

objectives: (*i*) maximize researchers' private utility, (*ii*) maximize total scientific output, or (*iii*) maximize social welfare given an assumption regarding the positive externalities of science. Using the model, we can solve for the allocation of inputs that achieves an objective while allowing for researchers' behavioral responses as they re-optimize their choices.[8] Notably, we can speak to changes in scientific output despite the fact that we never need to observe it.

Overall, we find evidence of a moderate degree of efficiency given our proposed objectives. The correlations between researchers' actual input levels and the optima implied by our model generally span 0.4–0.8; more productive researchers generally do have more inputs. However, there still appear to be significant gains from alternative allocations. Our counterfactuals suggest that total annual scientific output could be increased by approximately 160%. The private value to researchers of this growth is on the scale of 5% per year. The social value of this growth requires an assumption about the size of the positive externalities of science, which we explore.

To provide another way of characterizing the gains from reallocation, we ask the following: how much must funding levels increase under actual allocations in order to produce the same growth in output as our alternative allocations that hold input levels fixed? We find that total annual funding levels would need to increase roughly 40% ($155,000 per researcher per year) in order to achieve the same 160% growth in output our alternative allocations can achieve. Even conservative approaches to scaling these estimates up to the size of the population can easily imply gains from reallocation that are equivalent to multi-billion dollar increases in annual research funding.

Lastly, we unpack the counterfactuals to explore the following questions: Is the efficient allocation of inputs implied by the model more or less concentrated than the actual allocation? How does the reallocation of each input (i.e., money and time) independently change the results? How much does the efficient allocation change under different objectives? What is the distribution of input wedges (i.e., the difference between their optimal and actual input levels) across researchers? Is a researchers' input wedge predictable given their observable features? We also include a case study of gender differences in input wedges.

The remainder of the introduction provides a simplified motivating example of our methodology; Section 2 details the survey data; Section 3 describes our model of scientific production and consumption; Section 4 describes the survey experiments where we ask scientists to trade off their salary for inputs and reports the results of those experiments; Section 5 provides the productivity estimates; Section 6 contains our counterfactual allocation exercises and explorations thereof; and Section 7 concludes with a discussion about our results specifically and the potential usefulness of our methodology more generally.

[8] We use a number of constraints in order to keep our counterfactuals within the support of the data we observe and plausible in nature.

Overview of the Methodology

The principle of our methodology—that producers' input demand reveals information about their productivity—is a very old and pervasive concept. However, the particular way we make use of this principle is not. So, below we outline a simple example that illustrates our methodology and its usefulness, especially in analyses of sectors where output is not directly observable and inputs are not allocated via explicit price mechanisms.

Consider a set of profit-maximizing producers indexed by i who use a single input X they can obtain at non-zero heterogeneous costs $c_i(X_i)$ to produce output Q. For tractability, let the production technology be governed by a common functional form and heterogeneous productivity parameters α_i such that $Q_i = q(X_i; \alpha_i)$, where q is concave and well-behaved.[9] Demand for the output is a well-behaved homogeneous function of quantity $p(Q_i)$ common to all producers. Profits for each producer are given by price times quantity minus costs:

$$p\big(q(X_i;\alpha_i)\big)q(X_i;\alpha_i) - c_i(X_i) \ .$$

Producers' optimal input choices, X_i^*, clearly depend on input costs, output demand, and their productivity. At this point, the model is broadly applicable; these "profits" could in fact be measures of individuals' utility; the output of production could be virtually anything that consumers derive value from; input costs could be shadow prices, etc.

Now, we are interested in estimating the distribution of productivity, α_i, across producers, which is necessary to, among other things, make statements about allocative efficiency and policy counterfactuals. The vast majority of existing approaches to this task require data (or assumptions) on the quantities and prices of outputs. This is very difficult in many important settings such as the market for science.

To identify producers' productivity without observing output, consider the period after which producers have optimized and acquired their inputs, but before they have commenced production. At this time, producers are offered an opportunity to purchase a non-zero, fixed amount of additional inputs (e.g., at an auction); the amount of input being offered is given by X^+.Producers' willingness to pay for X^+ equals the additional profits (which, here, equal the additional revenues) they'd expect from using those inputs in production:

$$WTP_i(X^+) = p\big(q(X_i^* + X^+;\alpha_i)\big)q(X_i^* + X^+;\alpha_i) - p\big(q(X_i^*;\alpha_i)\big)q(X_i^*;\alpha_i) \ ,$$

which is independent of the input cost they faced in the market, and is greater than zero.

This expression is the key to our empirical approach—if you can observe or estimate the following objects: (1) producers' chosen input levels given current allocation mechanisms, X_i^*; (2) consumers' demand function, $p(Q_i)$; and (3) producers' willingness to pay for additional inputs, $WTP_i(X^+)$;

[9] $q(0;\alpha_i) = 0$, $q'(X_i;\alpha_i) > 0$, and $q''(X_i;\alpha_i) < 0 \ \forall \ i$.

then you can infer the productivity parameters α_i.

This approach does not require outputs to be observed, nor does it require inputs to be allocated via an explicit price mechanism—all it requires is that producers can price the value of the inputs based on their understanding of the demand for their output and their ability to convert inputs into that output productivity beliefs.

For example, consider a simple linear case that illustrates how the productivity term α_i can be written as a function of demand, initial input levels, and producers' willingness to pay for additional inputs:

$$q(X_i;\alpha_i) = \alpha_i X_i \text{ and } p(Q_i) = 1 - \varepsilon Q_i \implies \alpha_i = \frac{X^+ - WTP_i(X^+)}{2X_i^* X^+ \varepsilon + (X^+)^2 \varepsilon}.$$

The same logic applies to an alternative scenario where, after optimizing their input choices, but prior to actually using the inputs, producers were offered the opportunity to sell some amount of their inputs. Here, their willingness to accept would depend on their expected loss in profits, which in turn would depend on their productivity (and output demand, their initial input level, and the amount of inputs to be sold).

2 Survey and Data Overview

2.1 Survey Design

We use the National Survey of Academic Scientists (Myers et al. 2023) and provide a brief overview of the survey methodology here. The target population of the survey is US professors who conduct research at major institutions of higher education. To construct the sampling frame, information on professors was collected from the 158 largest institutions in the US by total R&D funding using the National Science Foundation's 2019 Higher Education R&D survey (HERD; National Science Foundation 2023).

The population consisted of 264,036 unique e-mails. A total of 131,672 individuals were e-mailed and 4,388 completed the survey (3.33%) completed the survey.[10] We then restrict the sample to the 4,003 individuals (91.2% of respondents) who reported being a professor, spending a non-zero amount of time on research, and having a non-zero salary from their primary institution.

During recruitment, incentives and reminders were randomly assigned. The four incentive arms were: (*i*) no incentive, (*ii*) entry into a lottery to win a gift card, (*iii*) the ability to vote for a set of charities to receive a donation, and (*iv*) both the second and third incentives. The reminder arms were zero, one, or two reminders. Each email was assigned one of the incentive arms and one of the reminder arms with equal probability, resulting in twelve possible incentive-reminder groups.

[10] The IRB approval permitted e-mailing only 50% of the population. The response rate is more than twice what has been obtained from sourcing academic researcher contacts from the corresponding author data contained within the publication record (e.g., Myers et al. 2020).

The randomized incentives and reminders provide us with instruments that we can use to implement a sample selection correction (Heckman 1979). The validity of this approach relies on having variables that cause entry into the sample (i.e., completing the survey) but does not affect the outcomes of interest. This allows us to correct for unobservable differences between the population and our sample. Appendix Table A1 reports the results from a regression of an indicator for survey completion on the different incentive and reminder arms, and it shows that all arms had a statistically significant positive effect on researchers' propensity to complete the survey.

In addition to correcting for unobservable differences, Myers et al. (2023) also checks the representativeness of the respondent sample by comparing it to the invited sample on observable characteristics. First, they explore a series of observable characteristics at the researcher-level by comparing the grant and publication histories of respondents and non-respondents using the Dimensions database (Digital Science 2018), which collects and disambiguates scientific metrics for researchers worldwide.[11] Appendix Figure A1 from Myers et al. (2023) shows invite-respondent overlap on various measures of scientific inputs and outputs. Overall, there is little difference in the respondents and non-respondents both economically and statistically speaking.

Looking at the institutional level, Appendix Figure A2 from Myers et al. (2023) shows invite-respondent overlap on various measures of institution funding derived from the HERD survey. As in the case of the researcher-level comparison, the distributions overlap to a large degree. Here, there is some statistically significant difference; on average, respondents come from institutions with slightly less funding (4–6%).

2.2 Summary Statistics: Researchers and their Inputs

Table 1 reports the mean, standard deviation, and median of the key covariates used in our analyses.

Fields of Study. Using the description of the professors department, we assign them first to a narrow set of twenty "minor" fields of study and then aggregate those fields into five broader "major" fields. Appendix Table A2 reports the distribution across the twenty minor fields. Unless otherwise noted, our counterfactual analyses will constrain the reallocation of inputs to only occur within the major fields.

Salaries and Research Inputs. The survey solicits a range of details regarding scientists' salary, their guaranteed funding (e.g., from prior grants or institutional guarantees), expected funds they will raise over the coming five years, and as their time allocations. Importantly, most variables are solicited as their expectations over the coming five years to ensure that this data spans the same time horizon as the thought experiments described below. In the Appendix, we replicate a test of respondents' self-reported salaries by comparing their self-reported salaries to the publicly-reported

[11] Using a fuzzy name matching process, Myers et al. (2023) are able to confidently match 87,000 (66%) of the researchers to their records in Dimensions.

salaries we are able to locate for a subset of researchers. Overall, there is a high degree of alignment (see Appendix Figure A3). Appendix Figure A4 plots full distributions of work time across each task, and Appendix Figure A5 plots full distributions of researching funding across the major fields for further detail. While there certainly is more research funding in traditionally capital-intensive fields (e.g., the natural sciences), there are a significant proportion of researchers in each major field with substantial research funding levels.

Position and Socio-demographics. Appendix Table A3 provides a full summary of all other major features collected regarding researchers' positions (e.g., rank and tenure status) and their socio-demographics (e.g., gender, race/ethnicity, citizenship). See Myers et al. (2023) for a more detailed investigation into these summary statistics.

2.3 Summary Statistics: Outputs

Our methodology for estimating productivity is explicitly designed to avoid the need to price or quantify the output produced by researchers' research efforts. Still, it is useful to have a more qualitative understanding of what researchers are producing. Thankfully, the survey includes a number of questions regarding researchers' output. Table 2 summarizes answers to these questions, which asked the frequency with which researchers intended to produce outputs of certain types (e.g., articles, books, materials, products) for certain types of audiences (e.g., peers, policymakers, businesses, general public) as well as other measures of riskiness and the degree to which the researcher focused on asking new questions or answering existing questions.

The traditional proxy for research output is peer-reviewed journal articles, and the data does indicate that journal articles written for academic peers is the most common output researchers intend to produce. However, there is considerable amount of variation and a significant amount of attention focused on producing output of types and for audiences that may never be codified in a journal article.

3 Model of Science

In this section, we present a model of researchers' utility from production and consumption. Subsection 3.1 presents the environment and optimal decisions. Subsection 3.2 describes how we incorporate externalities in production to define social welfare. Subsection 3.3 draws the connection between the model and the survey thought experiments that solicit researchers' willingness to pay (WTP) for different factors. Subsection 3.4 outlines our estimation approach. Additional details regarding the model are contained in Appendix B.

3.1 Researchers' Utility

The environment is populated by N researchers indexed by i whose primary choice is how much time to work (H_i) and how to allocate their total work between research R_i and fundraising F_i

over a 5-year horizon to maximize utility conditional on a contract from their primary institution, which is a triplet of state variables $\mathbf{S}_i = (M_i, G_i, D_i)$: salary M_i (\$), guaranteed funding G_i (\$), and administrative and teaching duties D_i (hours).[12] Researchers' indirect utility is given by:

$$\mathcal{V}(\mathbf{S}_i, \boldsymbol{\theta}_i, \boldsymbol{\mu}_i) = \max_{R_i, F_i, H_i} U_{1i}(M_i) + U_{2i}(Y_i) - U_{3i}(R_i, F_i, D_i) \tag{1a}$$

subject to

$$B_i = B_{\min} + G_i + \phi_i F_i \tag{1b}$$

$$R_i + F_i + D_i = H_i \tag{1c}$$

$$Y_i = \alpha_i B_i^{\gamma_i} R_i^{1-\gamma_i} \ , \tag{1d}$$

where $\boldsymbol{\theta}_i = (\alpha_i, \gamma_i, \phi_i)$ is the vector of individual-specific attributes related to scientific activity and $\boldsymbol{\mu}_i$ is a vector of parameters that govern the shape of the U functions. We assume that all salary is consumed and that research output Y_i is a Cobb-Douglas function of total funding B_i and research time R_i, with individual attributes governing productivity level and factor shares. Additionally, we impose the simple constraints that total funding amounts to a minimum $B_{\min}$ in addition to guaranteed funding and additional funds raised, and that all work hours are non-negative and bounded by a total hours endowment $H_{\max}$.

The functional forms in the utility function are as follows:

$$U_{1i}(\cdot) = \omega \frac{(M_i)^{1-\sigma_i}}{1-\sigma_i} \tag{2a}$$

$$U_{2i}(\cdot) = \frac{(Y_i)^{1-\eta_i}}{1-\eta_i} \tag{2b}$$

$$U_{3i}(\cdot) = \psi \frac{(R_i + F_i + D_i^{\xi_i})^{1+\zeta_i}}{1+\zeta_i} \ . \tag{2c}$$

U_{1i} describes the utility from salary and is governed by a common slope parameter ($\omega > 0$) and a researcher-specific curvature parameter $\sigma_i > 0$. U_{2i} describes the utility from scientific output. It is effectively a revenue function whose shape depends on the market's demand for a researchers' output, and it is governed by a researcher-specific curvature parameter $\eta_i \in (0, 1)$. U_{3i} describes the disutility from efforts across different tasks, and it is governed by a common slope parameter ($\psi > 0$) and two researcher-specific curvature parameters ($\xi_i > 0, \zeta_i > 0$).[13] Thus, the vector $\boldsymbol{\mu}_i$ collects the parameters $(\omega, \sigma_i, \eta_i, \psi, \xi_i, \zeta_i)$.

The policy functions $\mathcal{R}(\mathbf{S}_i, \boldsymbol{\theta}_i, \boldsymbol{\mu}_i)$, $\mathcal{F}(\cdot)$, and $\mathcal{H}(\cdot)$ characterize the solutions to Equation (1). For each individual researcher, they determine optimal (R_i^*, F_i^*, H_i^*) as a function of states, attributes, and parameters. The derivations of these policy functions are described in Appendix B.1.

[12] We're using the word contract loosely, since, for example, a researchers' guaranteed funding may come in part from future flows of funding from grants obtained outside the institution.

[13] The ξ_i parameter allows for additional disutility from duty-related work (e.g., administration or teaching), which we found to be a parameter whose introduction improved the model's fit of the data.

Ideally, we would have enough variation to estimate all researcher-specific attributes including the production-related parameters $(\alpha_i, \gamma_i, \phi_i)$ as well as all of the consumption-related parameters that govern researchers' utility from salary, output, and effort $(\sigma_i, \eta_i, \xi_i, \zeta_i)$. However, as will be clear below, we only have enough variation in the survey experiments to uniquely identify the production-related parameters $(\alpha_i, \gamma_i, \phi_i)$. Thus, we choose to specify each of the consumption-related parameters to be a parametric function of researchers' observable features.

Fortunately, we have a large vector of observable features $(\mathbf{X}_i)$ that describe researchers' positions, their backgrounds, and their subjective descriptions of their scientific output (i.e., the features summarized in Table 2 and Appendix Table A3). This is useful because it can allow us to incorporate more heterogeneity into the consumption components of the model and limit the degree to which variation in the data might otherwise cause us to overstate the heterogeneity in the production components of the model.

Unfortunately, the computational demands of estimating the model limit the flexibility with which we can incorporate the dozens of features available. Thus, to balance the benefits of allowing for heterogeneity in consumption with the benefits of simpler estimation, we use Principal Component Analysis (PCA) to reduce the full set of observable features $(\mathbf{X}_i)$ into a one-dimensional index. We refer to this resulting index as researchers' "type," (T_i) and we specify each of the consumption-related parameters $(\sigma_i, \eta_i, \xi_i, \zeta_i)$ to be simple, univariate functions of each researchers' type and homogeneous, deep parameters. Specifically, the parameter that governs the utility from private consumption is parameterized as: $\sigma_i = \exp(\delta_{\sigma,0} + T_i\delta_{\sigma,1})$. We similarly parameterize ζ_i and ξ_i with exponential functions with intercept and slope $(\delta_{\zeta,0}, \delta_{\zeta,1})$ and $(\delta_{\xi,0}, \delta_{\xi,1})$, respectively. The curvature of utility in scientific output, η_i, is modeled with a logistic function $[1 + \exp(\delta_{\eta,0} + T_i\delta_{\eta,1})]^{-1}$ bounded in the interval $(0, 1)$. Therefore, we express individual-specific parameters $\boldsymbol{\mu}_i(T_i, \tilde{\boldsymbol{\mu}})$ as a function of researcher's type T_i and of the vector of common parameters to be estimated, denoted by $\tilde{\boldsymbol{\mu}} = (\omega, \psi, \boldsymbol{\delta})$, with $\boldsymbol{\delta} = (\delta_{\sigma,0}, \delta_{\sigma,1}, \delta_{\zeta,0}, \delta_{\zeta,1}, \delta_{\xi,0}, \delta_{\xi,1}, \delta_{\eta,0}, \delta_{\eta,1})$.

Allowing the $\boldsymbol{\mu}_i$ parameters to be type-specific is not a panacea, but it helps reduce the degree to which variation in the experimental data that is truly driven by heterogeneous preferences or heterogeneous demand for scientific output contaminates our productivity estimates. We further detail the estimation process below.

3.2 Social Value

To incorporate the externalities of scientific production (e.g., Nelson 1959; Arrow 1962), which we assume to be net positive, we define the social value produced by each researcher as:

$$\mathcal{W}(\mathbf{S}_i, \boldsymbol{\theta}_i, \boldsymbol{\mu}_i, \kappa) = U_{1i}(M_i) + \kappa U_{2i}(Y_i^*) - U_{3i}(R_i^*, F_i^*, D_i) \,, \tag{3}$$

where researchers' privately optimal choices and output are given by R_i^*, F_i^* and Y_i^*. κ reflects the size of the externalities associated with researchers' output. Thus, $1/\kappa$ gives the share of the social

value generated by each researchers' output that they themselves capture and $1 - 1/\kappa$ is the size of the positive externality.

This formulation of externalities is an admittedly ad hoc way of implicitly modeling consumer surplus as some constant multiple of producer surplus. However, it has a clear economic interpretation as a measure of appropriation, and we can draw on prior studies that have sought to identify precisely this measure. Studies that focus on commercial innovators have found producers' value capture to be as large as 15% ($\kappa = 6.7$) and as low as 2% ($\kappa = 50$) (Nordhaus 2004; Philipson and Jena 2006; Jena and Philipson 2008; Lakdawalla et al. 2010). In our empirical analyses, our baseline assumption is $\kappa = 10$, which implies researchers capture 10% of the social value they create, and we also explore some scenarios with an assumption of $\kappa = 100$ (social value capture equals 1%).[14]

3.3 Input Demand and Productivity

This subsection illustrates how our model can be used to generate predictions about a researcher's willingness to pay (or accept) for more (or less) inputs than what existing allocation schemes would lead them to obtain.

Consider the period prior to production but after which researchers have formed their expectations about their time allocations and, therefore, their expectations about their utility over the 5-year horizon. Now, researchers are offered some alternative contracts by their primary institution that vary either funding guarantees ($G \rightarrow \widetilde{G}$) or administrative duties ($D \rightarrow \widetilde{D}$), but leave the salary unspecified.

For each offer indexed by j and characterized by $(\widetilde{G}_{ij}, \widetilde{D}_{ij})$, there is some salary $(\widehat{M}_{ij})$ that would make the researcher indifferent between their actual contract (M_i, G_i, D_i) and the offer such that:

$$V_i \equiv \mathcal{V}(M_i, G_i, D_i, \boldsymbol{\theta}_i, ...) = \mathcal{V}(\widehat{M}_{ij}, \widetilde{G}_{ij}, \widetilde{D}_{ij}, \boldsymbol{\theta}_i, ...) \; . \tag{4}$$

Importantly, $\mathcal{V}(\cdot)$ incorporates behavioral responses such that

$$\underbrace{\mathcal{R}(\mathbf{S}_i, \boldsymbol{\theta}_i, ...)}_{\substack{\text{expected research} \\ \text{time in actual} \\ \text{position}}} \gtreqless \underbrace{\mathcal{R}(\widetilde{\mathbf{S}}_{ij}, \boldsymbol{\theta}_i, ...)}_{\substack{\text{expected research} \\ \text{time in alternative} \\ \text{position}}}$$

where and likewise for $\mathcal{F}(\cdot)$ and $\mathcal{H}(\cdot)$; researchers evaluate their net indirect utility in the alternative position after they will have re-optimized their time allocations given the new state variables.[15]

For example, if a researcher is offered more guaranteed funding ($G_i < \widetilde{G}_{ij}$) then they will forecast how this additional funding will lead to changes in their optimal time allocations, their expected

[14] Note also that this approach implies that researchers' private utility reflects the case where $\kappa = 1$.

[15] Or put alternatively, $\widetilde{R}^*_{ij} \gtreqless R^*_i$, $\widetilde{F}^*_{ij} \gtreqless F^*_i$, and $\widetilde{H}^*_{ij} \gtreqless H^*_i$.

input levels, their expected output levels (per their productivity beliefs!), and their expected indirect utility. The total increase in their expected indirect utility can be priced by the researcher and stated as a new, lower salary ($M_i > \widehat{M}_{ij}$).

Importantly, we can write that alternative salary ($\widehat{M}_{ij}$) that makes researchers indifferent between their actual position and the offer to be some function of known variables and the unknown parameters to be estimated:

$$\widehat{M}_{ij} = \Pi(\widetilde{G}_{ij}, \widetilde{D}_{ij}, M_i, G_i, D_i, \boldsymbol{\theta}_i, ...) \ , \tag{5}$$

where Π is determined by functional forms of, and optimality conditions implied by, Equation (1). Thus, if we know the left-hand side of Equation (5) from some experimental offers, and we also know all components of Π except for the parameters, then we can estimate those parameters.

3.4 Estimation

We infer individual attributes $\boldsymbol{\theta}_i$ and parameters $\tilde{\boldsymbol{\mu}}$ jointly, using a mix of calibration and estimation. First, we set minimum funding $B_{min} = \$5,000$ and fix total hours endowment $\bar{H}$ to 62 work hours per week (approximately the 90th percentile of the observed work hours). Second, we estimate individual attributes and common parameters of the utility function—including deep parameters that determine individual-specific parameters in $\boldsymbol{\mu}_i$ as functions of researcher's type T_i—mapping theoretical objects in the model to their theoretical counterparts. Specifically, we use survey information on declared hours worked, research time, fundraising time, guaranteed funding, expected additional funds raised, duties, and the salaries reported in the alternative offer experiments, which we map to model counterparts $(H_i, R_i, F_i, G_i, \phi_i F_i, D_i, \{M_{ij}\}_{j=1}^4)$, respectively.

Appendix B contains the full details on our estimation routine, which we briefly outline here. Conditional on some common parameters, we can infer individual-specific attributes $\boldsymbol{\theta}_i$ from researchers' optimality conditions and model's structure. Therefore, we first describe this mapping, and then we proceed to present estimation of the common parameters. One important note is that we have separate estimation strategies for inferring individual attributes $\boldsymbol{\theta}_i$ among researchers with non-zero fundraising time ($F_i > 0$) and those at a corner solution ($F_i = 0$).

3.4.1 Inference on Individual Attributes given Parameters

For the former group, we first infer fundraising ability ϕ_i by exploiting the identity equating the observed, expected additional funding (EG_i) to $\phi_i F_i$. Therefore, $\forall i = 1, ..., N_F$:

$$\widehat{\phi}_i = \frac{EG_i}{F_i} \ . \tag{6}$$

Expression (6) is well-defined if and only if observed $F_i > 0$, which motivates why the inference strategy must differ for scientists with $F_i = 0$ at observed allocations.[16] As a second step, con-

[16] For numerical stability, we constrain $\widehat{\phi}_i \geq 1$.

ditional on $\widehat{\phi}_i$, we infer factor shares $\widehat{\gamma}_i$ from scientists' first order condition in fundraising time (B3), which we derive in Appendix B.1. For the chosen functional forms, this takes the intuitive expression:

$$\widehat{\gamma}_i = \frac{B_i}{B_i + \phi_i R_i} = \frac{B_{\min} + G_i + \widehat{\phi}_i F_i}{B_{\min} + G_i + \widehat{\phi}_i F_i + \widehat{\phi}_i R_i} \,. \tag{7}$$

Equation (7) states that γ_i constitutes the weight of total funding over the total dollar-value of inputs used in scientific activity, with the last term of the denominator being the dollar-valued opportunity cost of research time relative to fundraising.

Finally, the optimality condition for total hours worked—Equation (B3)—determines productivity $\widehat{\alpha}_i$ as a function of parameters, observed allocations, and $\widehat{\phi}_i$ and $\widehat{\gamma}_i$:

$$\eta_i(1-\widehat{\gamma}_i)\widehat{\alpha}_i^{1-\eta_i}(B_{\min} + G_i + \widehat{\phi}_i F_i)^{(1-\eta_i)\widehat{\gamma}_i} R_i^{(1-\eta_i)(1-\widehat{\gamma}_i)-1} = \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} \,, \tag{8}$$

This equation holds exactly if $H_i < H_{\max}$. Therefore, Equations (6), (7), and (8) determine individual attributes as functions of parameters and observed allocations for researchers with positive fundraising time.

For the second group of scientists with zero time allocated to fundraising activity, unfortunately we can neither infer $\widehat{\phi}_i$ from Equation (6), nor we can compute funding weight $\widehat{\gamma}_i$ based on optimality condition (B3).[17]

Therefore, we assume that ϕ_i and γ_i are parametric polynomial functions of state variables and we estimate such functions based on the former sub-sample of scientists for which observed $F_i > 0$. Specifically, we consider the following specifications:

$$\phi_i = \exp\left\{\sum_{p=1}^{3} \beta_{G,p} G_i^p + \sum_{p=1}^{3} \beta_{D,p} D_i^p + \sum_{p=1}^{3} \beta_{M,p} M_i^p\right\} \tag{9}$$

$$\gamma_i = \left(1 + \exp\left\{\sum_{p=1}^{3} \iota_{G,p} G_i^p + \sum_{p=1}^{3} \iota_{D,p} D_i^p + \sum_{p=1}^{3} \iota_{M,p} M_i^p\right\}\right)^{-1} \,, \tag{10}$$

which we estimate non-linearly as a Poisson model and a logistic regression. We thus recover $\widehat{\phi}_i$ and $\widehat{\gamma}_i$ for the second sub-sample of researchers with no fundraising activity as the fitted values of models (9) and (10), respectively.[18]

Finally, given the estimates attributes $\widehat{\phi}_i$ and $\widehat{\gamma}_i$, the state variables, and the vector of common parameters $\tilde{\boldsymbol{\mu}}$, we structurally infer productivity beliefs $\widehat{\alpha}_i$ through equation (8).

[17] The Lagrange multiplier $\lambda_{F,i}$ is strictly positive and unknown.

[18] Whenever the fitted $(\widehat{\phi}_i, \widehat{\gamma}_i)$ combination would imply, given the observed state variables and for a specific vector of parameters $\boldsymbol{\mu}_i$, that the optimal (observed) time allocation to fundraising would be strictly positive, we re-scale $\widehat{\phi}_i$ below the individual-specific lower bound below which the optimal fundraising time at observed states is indeed null.

3.4.2 Estimation of Common Parameters

Given estimates of individual attributes $\widehat{\boldsymbol{\theta}}_i = (\widehat{\alpha}_i, \widehat{\gamma}_i, \widehat{\phi}_i)$ as functions of parameters $\tilde{\boldsymbol{\mu}} = (\omega, \psi, \boldsymbol{\delta})$, calibrated values $(B_{\min}, H_{\max})$, and observed allocations, we estimate the vector $\tilde{\boldsymbol{\mu}}$ by generalized method of moments as:

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\tilde{\boldsymbol{\mu}}} \mathcal{L}(\tilde{\boldsymbol{\mu}}) = \arg\min_{\tilde{\boldsymbol{\mu}}} \sum_{i=1}^{N} \sum_{j=1}^{4} \left(M_{ij} - \widehat{M}_{ij}(\mathbf{S}_{ij}, \widehat{\boldsymbol{\theta}}_i(\boldsymbol{\mu}_i(\tilde{\boldsymbol{\mu}}, T_i)), \boldsymbol{\mu}_i(\tilde{\boldsymbol{\mu}}, T_i)) \right)^2 , \tag{11}$$

where M_{ij} is researcher i's answer to thought experiment j in the data and $\widehat{M}_{ij}$ is its model-based counterpart, which is in turn am implicit function of the vector of counterfactual states $\mathbf{S}_i$, researcher's type T_i, and of the estimand $\tilde{\boldsymbol{\mu}}_i$ which determines individual attributes $\widehat{\boldsymbol{\theta}}_i$ and individual-specific parameters $\boldsymbol{\mu}_i$.[19] Parameter estimates $\widehat{\boldsymbol{\mu}}$ solve (11) conditional on parameter restrictions specified in Section 3.1. Appendix B.2 provides additional details on the search algorithm.

4 Survey Experiment

Here, we describe the thought experiments in the survey and how they connect to the model. Then we report the distribution of responses and some tests of face validity and other biases (e.g., sample selection).

4.1 Soliciting Willingness to Pay

WTP for Funding and Time. Respondents are presented with four hypothetical scenarios. They are asked to imagine that their primary institution has offered them: (1) an increase of $250,000 in guaranteed funding in exchange for a lower salary, (2) an increase of $1,000,000 in guaranteed funding in exchange for a lower salary, (3) an elimination of all administrative duties in exchange for a lower salary, and (4) an increase in duties by 20 hours per month over a 5-year period in exchange for a higher salary.[20] All of these hypotheticals are posed over a 5-year span to fix the time horizon for all respondents (and is why the model below is described as capturing a 5-year span). In order to solicit the salary at which they are indifferent between the current position and each offer, respondents are asked to report the lowest salary they would be willing to accept the offer.[21] Importantly, the survey only solicits researchers' WTP for more funding. This is important because, when we reallocate funding, often we are taking funding away from researchers. Thus, we must assume that the value of infra-marginal funding is equal to the value of marginal funding (after filtered through the model of the production function). In the case of time, the survey solicits

[19] To save notation, we omit that $\widehat{M}_{ij}$ are also a function of calibrated $(B_{\min}, H_{\max})$.

[20] Researchers who report no administrative duties are not shown (3).

[21] Pilot tests indicated that researchers could more easily report the lowest salary that could make them take the offer as opposed to the literal salary at indifference, and since indifference is an vanishingly small amount less than this reported value, we treat their answer as the amount at indifference

both WTP (for more free time) and willingness to accept (WTA; for less free time) and we will be treating those as symmetric (again, after filtered through the model of the production function). Appendix Figure A6 displays examples of how these thought experiments appeared to the survey respondents.

WTP for a Benchmark Good. When soliciting willingness to pay, especially in an un-incentivized manner, it is always possible that respondents under- or overstate their price sensitivity. In order to test for this, we follow Dizon-Ross and Jayachandran (2022) and explore respondents willingness to pay for a "benchmark good." Dizon-Ross and Jayachandran (2022) note that, if you solicit an subject's willingness to pay for a good whose value (to the subject) is plausibly uncorrelated with the value of the focal good, then any correlation between the two willingness to pays reported can be (partially) attributed to systematic noise. The survey asks respondents to report the most amount of money they would be willing to pay per month for high-speed internet access at their primary residence, which we use as our benchmark good since it is plausibly uncorrelated with a researchers' productivity but of non-zero value to most respondents. In Appendix Table A4, we show that respondents' stated WTP for research funding and free time are rarely correlated with their stated WTP, which suggests a small amount of variation in researchers' answers is attributable to systematic noise.

4.2 Connection to Model: WTP and Compensating Variation

Four survey thought experiments elicit the compensating variation of individual researchers in relation to (i) an increase of \$250,000 in guaranteed funding G_i, (ii) an increase of \$1,000,000 in guaranteed funding G_i, (iii) a reduction of duties D_i to 0, and (iv) an increase in duties by 20 hours per month over a 5-year period. In other words, we ask for individual income levels $\widehat{M}_{ij}$, where $j = ,1,2,3,4$ indexes the four thought experiments, that would make researcher's utility in the counterfactual scenario equal to her indirect utility V_i^* at current allocations. Formally, counterfactual guaranteed funding and duties in the four experiments are:

$$
\begin{aligned}
(\widetilde{G}_{i1}\,,\widetilde{D}_{i1}) &= \big((G_i + \$250{,}000)\,, D_i\big) \\
(\widetilde{G}_{i2}\,,\widetilde{D}_{i2}) &= \big((G_i + \$1{,}000{,}000)\,, D_i\big) \\
(\widetilde{G}_{i3}\,,\widetilde{D}_{i3}) &= \big(G_i\,, 0\big) \\
(\widetilde{G}_{i4}\,,\widetilde{D}_{i4}) &= \big(G_i\,, (D_i + 20 \text{ hours/month})\big)\,,
\end{aligned}
$$

where G_i and D_i represent actual states.

As noted above, the values of $\widehat{M}_{ij}$ reported in these four experiments make the researcher indifferent between all possibilities (per Eq. 4), and therefore we can write these reported values to be a known function of observable data and the unknown parameters to be estimated (per Eq. 5).

4.3 Survey Experiment Results

Figure 1 plots the distribution of researchers' WTP for guaranteed research funding and free time (in the form of WTP for less duties or WTA for more duties). The values shown are averaged over the two thought experiments for either factor and converted to a per-dollar basis. As evidenced by Figure 1, there is considerable variation in WTP responses across researchers. Some of this variation reflects different preferences and constraints, but some also reflects heterogeneous productivity across researchers. The model described above allows us to separate these two forces.

Testing the validity of these WTP responses is a difficult but important task. They are the key ingredient of our entire exercise, but they may be driven in part by sample selection effects and/or behavioral biases and/or noise due to a simple lack of effort on the part of respondents. However, despite reflecting very practical trade-offs that professors are posed with throughout their careers, systematic data has not been collected in this way before to our knowledge. Thus, it is not clear what "plausible" variation would look like here. Still, we can conduct some tests motivated by economic and statistical theory to explore how reasonable these distributions of WTP are.

Motivated by the notion of opportunity costs, we test how WTP varies with researchers' implied hourly wage (per their annual salary divided by their total hours of work). Assuming this implied wage rate is a proxy for researchers' opportunity costs, it should be the case that researchers with higher hourly wages are willing to pay more for their free time. In Appendix Figure A7a, we see that is indeed the case. Furthermore, it should be the case that, conditional on opportunity costs of time, researchers who expect to spend a larger amount of time on fundraising should be more willing to pay for guaranteed funding. In Appendix Figure A7b, we see that is indeed also the case.

As another test of the WTP estimates, Appendix Tables A4 and A5 report a series of analyses where we regress researchers' answers to the thought experiments on the state and choice variables of the model, and possibly also (*i*) the large vector of observable features describing each researcher ($\mathbf{X}_i$), (*ii*) an inverse Mills ratio constructed using the randomized survey participation incentives following Heckman (1979), and/or (*iii*) the researchers' WTP for the benchmark good (high-speed internet at their home) following Dizon-Ross and Jayachandran (2022).

We find the variables of the model can explain a large share (80–90%) of the variation in WTP responses (see Appendix Table A4). This supports the model's ingredients as being key determinants of researchers' answers. Secondly, we find that the inverse Mills ratio is not a statistically significant predictor of responses (see Appendix Table A4). Under the assumptions outlined in Heckman (1979), this supports the notion that respondents did not differentially select into our sample as a function of their WTP for inputs and suggests some generalizability of our responses. Finally, we find that researchers' WTP for the benchmark good (high-speed internet at their house) is correlated with their WTP for inputs. Under the assumption that researchers' demand for home internet is truly uncorrelated with their productivity, this provides some evidence of systematic

noise in how respondents are reporting their WTP in these experiments. But, the magnitudes of this relationship are quite small and the benchmark good WTP is included in our vector of features we use to make the researcher type index (that determines heterogeneity in the consumption parameters), which gives us some confidence that this systematic noise is small and the model will adjust for it to some degree. Overall, despite being un-incentivized experiments, researchers' responses appear to be behave as expected and be of plausible magnitudes.

5 Estimates for the Model of Science

5.1 A View of the Utility Function

The parameter estimates from the for the model are reported in Appendix Table C1. Appendix Figure C1 shows model fit by reporting the distribution of experimental answers, the model's predictions, and the model's errors.

To illustrate the utility function in a more intuitive way, Figure 2 shows how an average researcher's utility depends on the levels of four key variables: salary, administrative duties, research budget, and research time. The Figure shows the percent change in a researcher's utility levels as the variable is increased from the 10^{th} percentile level to the 90^{th} percentile level while all other variables and parameters are held fixed at the sample averages.[22]

In terms of magnitude, salary and administrative duties are the most important to researchers' utility. Shifting the average researcher's salary from the 10^{th} percentile to the 90^{th} percentile increases their utility by roughly 60%. Shifting administrative duties the same relative magnitude decreases their utility by roughly 20%. Relatively equivalent increases in the inputs to the scientific production function (research budget and time) increase researchers' utility by only a few percentage points. This is partly because the payoff to additional scientific output is very low relative to salary, and, in the case of research time, because there is some disutility from the additional work hours that offsets some of the positive utility from the resulting output.

5.2 Productivity Distributions

Figure 3 shows the unconditional distribution of individual-specific production function parameters γ_i (funding intensity) and α_i (productivity). Figure 3a displays the distribution of the γ_i parameter, which describes the relative weight of funding versus research time in researchers' production functions. On average, funding accounts for nearly 40% of the optimal input mix.The distribution highlights a significant degree of heterogeneity across researchers, with roughly 20% of our sample having a funding intensity either larger than 0.6 or smaller than 0.2.

As both an interesting exercise and test of face validity, Appendix Figure C2 plots the average funding intensity parameters (γ_i) for the twenty minor fields represented in the sample. As one

[22] For simplicity, behavioral responses where researchers re-optimize are not incorporated into Figure 2.

would expect, γ_i is highest in chemistry and engineering, where research is often capital intensive and involves the use of expensive lab equipment. In contrast, and as expected, the social sciences (e.g., economics, political science) display the lowest funding intensity on average. Still, Figure C2 provides another view of the heterogeneity in funding intensity across researchers even within these narrower fields of study.

Figure 3b displays the distribution of the α_i parameter, which is our productivity metric analogous to TFP. Consistent with existing evidence based on indirect proxies for scientific productivity (e.g., citations; Seglen 1992), our estimates reveal a large skew in researchers' beliefs about their productivity.

As a test of face validity, Append Table C2 reports regression results that test for associations between researchers' productivity (α_i) and their performance per traditional metrics of scientific productivity, which are often more simply just output levels (e.g., publications, citations). Under some reasonable assumptions, we expect a non-zero association between these metrics, and we find it. Researchers with higher productivity beliefs are more likely to have more actual publication output whether that output is measured using recent publication counts or citation-weighted counts. This provides a signal that our estimates do in fact reflect real productivity differences across researchers.

To get some sense as to the (private) value of these productivity differences, Appendix Figure C3 shows the results of an exercise where we solve for researchers' WTP for productivity differences based on their willingness to trade utility from salary for utility from productivity. This allows for a comparison in dollar equivalents. Per this conversion, the heterogeneity in productivity appears economically meaningful to researchers, who are often willing to pay many thousands of dollars per year to be more productive.

But while Figure 3b shows a very large degree of heterogeneity, some fraction of this variation may be due to differences in the demand for their output or the nature of science within their fields of study. It is for an analogous reason that most studies on the industrial organization of firms report productivity distributions based only on within-industry variation. Thus, Figure 4 shows the distribution of researcher productivity (α_i) after various controls for field-specific (or "industry-specific") variation are introduced.

First, Figure 4 shows the raw distribution of productivity levels, which mirrors the distribution shown in Figure 3b but now on a logarithmic scale. Next, we introduce a first set of controls for field-specific variation in production in demand: a set of major-field fixed effects (*Field-FE*), and the one-dimension PCA-based type index variable ($\mathbf{X}$-*index*); and we report the distribution of residual productivity levels. Finally, we introduce additional controls: the researchers' specific values of funding intensity (γ) and fundraising efficiency (ϕ). The purpose of these controls are to focus only on variation in the α estimates that is more plausibly driven by actual differences in researcher productivity and less likely to reflect idiosyncratic preferences or demand.

Figure 4 also reports ratio of the 90^{th} and 10^{th} percentiles, the "90–10 ratio" measure of variance commonly reported in traditional, firm-level studies. The 90–10 ratio of the raw productivity estimates is dramatic, but, even after conditioning on the field-specific controls, we still find a large amount of variation. The 90^{th} percentile researcher believes they are roughly 6–12 times as productive as the 10^{th} percentile researcher.

It is difficult to benchmark this distribution, but firm-level estimates from manufacturing sectors around the world typically find 90–10 ratios on the scale of 2:1 to 5:1 (Syverson 2011). Thus, the degree of dispersion we find suggests that the scope for misallocation is substantial, which we investigate next.

6 Counterfactuals and Allocative Efficiency

6.1 Overview

In this section, we use our estimates of researchers' productivities to analyze the allocative efficiency of scientific inputs through the lens of our model. Specifically, we consider guaranteed funding (G) and duties (D) as policy levers that a planner can use to maximize some objective subject to some constraints. Importantly, we search for the allocations of guaranteed funding and duties the maximize an objective after accounting for researchers' endogenous behavioral responses to the planner's decisions. The two margins for behavioral responses are time spent fundraising (to obtain additional funds) and time spent on research (to directly produce output). Of course in reality, researchers may endogenously adjust to the reallocation of funds and duties along many other margins (e.g., which researchers they collaborate with, the types of projects they pursue). These complexities are beyond the scope of our model, and we believe that adjustments to time allocations is one of the most first-order responses and it allows us to keep the problem tractable.

We study optimal allocations under three different objectives. First, we consider an utilitarian objective of maximizing researchers' aggregate utility. This exercise provides a useful benchmark to think about misallocation within the context of our model and the gains to be had simply by focusing on the utility of researchers. Second, we ask which allocations of funding and duties would maximize social value assuming that researchers capture 10% of the social value that their marginal output produces ($\kappa = 10$). This exercise represents a experiment that more explicitly incorporates externalities in researchers' production—the social value they create that they do not internalize. Lastly, we search for the allocation that maximizes total output. And to aid in the interpretation of the results from these counterfactuals, we conduct an exercise where we ask how much total research funding must be increased *using actual allocations* to achieve the same growth in scientific output that we are able to achieve *using actual funding levels* in alternative allocations.

Importantly, the presence of externalities looms large here. As noted previously, our approach implicitly assumes that all researchers' production involves the same amount of externalities (per the common parameter κ). In practice, these externalities may vary greatly across fields, which is

why we mostly only consider reallocations *within* the five major fields of researchers in our sample (i.e., Engineering, math and related; Humanities and related; Medical and health sciences; Natural sciences; Social sciences). Beyond this, we explore a range of alternative constraints on how inputs are reallocated to draw broader conclusions about the gains from reallocation here. Appendix B contains further details on how we specify the optimization problems.

6.2 Summary of Results

Table 3 provides a broad summary of our findings. As a baseline, the first column reports summary statistics for research inputs given current allocations. We do not report the levels of output and utility given current allocations since the units of these metrics are not objectively interpretable.

Columns (2–7) of Table 3 report the percentage point changes in research inputs, outputs, and value under the six main counterfactuals we explore. The values represent the equilibrium change in levels after accounting for researchers' behavioral responses. Furthermore, the bottom two rows of these columns (under *Input reallocation*) report the amount of the input reallocated in percentage terms of the total amount of input in the sample (e.g., if there was $100 million in total research funding, a value of 50% indicates that the optimal policy would reallocate $50 million).

The other columns report results from the counterfactuals that reallocate duties (ΔD) guaranteed funding (ΔG), or both ($\Delta D, G$) in order to achieve the objective of maximizing scientists' aggregate utility (U), aggregate social value (W), or total output (Y) while holding the total amount of funding in each major field fixed, or allowing it grow as long as scientists are willing to commit time to fundraising (*Unconstrained* B). In all cases, social value is evaluated where $\kappa = 10$ (implying that scientists capture 10% of the value of their marginal output).

An initial result that Columns 2–3 of Table 3 highlights is the following: reallocating researchers' time constraints (i.e., their administrative duties is the key lever for affecting their utility while reallocating their budgets is the key lever for affecting their output. And to a first approximation, the aggregate changes from jointly optimizing these two factors is roughly equivalent to the sum of independently optimizing them.

Broadly speaking, the model suggests that there are very large gains in output to be had from alternative allocations of research budgets. Whether inputs are allocated to maximize private value, social value, or total output, the new allocations yield roughly 140–160% more output. This growth in output (and the behavioral changes in time allocations) leads to significant gains in value on the scale of 5% for researchers' private utility and 15% for social welfare (given our assumption about the size of externalities). As these are annual flow metrics of value, these magnitudes appear to be meaningful when compared to traditional scales of meaningful macroeconomic growth rates.

However, it is challenging to get a subjective sense of the magnitudes of the gains from reallocation reported in Table 3. Thus, we estimate how much more funding under current allocations would be

necessary to achieve the same growth in scientific output that our counterfactual allocations achieve using current funding levels. Table 4 reports these estimates under both the "*Mechanical*" scenario that does not incorporate researchers' behavioral response and one that does ("*Behavioral*").

This exercise indicates that research budgets would need to increase roughly 40% for the actual allocations to achieve the same growth in output (approximately 160%) that we observe in most of the counterfactuals. On an annual basis this amounts to roughly $160,000 per researcher. In aggregate, this sums to roughly $240 million per year in our sample. Under the assumption of insignificant non-response bias, this implies that the gains from a more efficient allocation are equivalent to roughly $14 billion per year if we aggregated across the entire population of researchers that were targeted in the survey.

As evidenced by Table 3 Columns 4–6, optimizing to maximize researchers' aggregate utility (U), social value (W), or total output (Y) all yield qualitatively similar outcomes. This is primarily driven by the fact that researchers have control over their time. Whatever the planner hopes to maximize, the general approach is to ensure that the most productive researchers are incentivized to spend more time working on their science.[23]

As a relatively implausible but useful comparison, Table 3 Column 7 shows the result of a scenario where the total research budget is unconstrained—researchers can fundraise as much money as they would like (while still facing the disutility of fundraising effort and the constraints on their total work time). Compared to the benchmark reallocation scenario reported in Column 4, this scenario of an unlimited budget (Col. 7) actually yields very little additional gains in output and value. The diminishing returns of funding in the scientific production function and the disutility and constraints of work time limit the equilibrium growth in the total budget to be only roughly 14%, and only a few more percentage points of output and a negligible amount of value is produced with these marginal funds. While a full cost-benefit analyses of these investments in science is beyond the scope of this paper, this result suggests that the total amount of funding in this economy is likely not far below the optimum because so little is left to gain from additional investment (conditional on those investments being optimally allocated).

Additional counterfactuals using alternative constraints and specifications are reported in Appendix Table C3. One of the most important results shown there is that allowing for reallocation across major fields does not yield much additional output or value compared to the scenarios where reallocation is constrained to occur only within major fields (as is done in the counterfactuals reported in Table 3). This finding suggests that the degree of misallocation within fields is as large (or larger) than it is across fields.

[23] In the case of maximizing private versus social welfare, this is also partly a result of the assumption that the positive externalities of output are a constant multiple (per κ).

6.3 Input Wedges: Actual Versus Optimal

The results summarized in Table 3 suggest that reallocating researchers' time constraints and funding can yield significant gains in output and value. In order to better understand how these gains are achieved, we now focus on a single counterfactual specification and compare the actual and optimized input levels. For simplicity, we focus on the scenario where duties and funding are jointly optimized to maximize researchers' utility, which corresponds to the counterfactual described in Column (4) of Table 3.

Figure 5 shows the correlation between observed and optimal research hours (5a) and total funding (5b), which is indicative of the degree of misallocation. In both cases, there is clearly a positive correlation suggesting that inputs are allocated efficiently to some degree. The optimal input levels describe 16% of the variation in actual research time and 57% of the variation in actual research funding. Loosely speaking, research time appears to be more misallocated than research funding given our assumed objective.

Appendix Figure C4 illustrates the same relationship between actual and optimal inputs, but instead focuses on the policy levers which are allocated pre-behavioral response (i.e., administrative duties and guaranteed research funding). There we find effectively no relationship between actual and optimal levels. Thus, while the optimum implied by the model is an equilibrium allocation of inputs that is relatively similarly to what is actually observed, the model suggests a very different way of achieving it.

Appendix Figure C5 shows the distribution of actual and optimal levels first for the two policy levers (*Administrative Duties* and *Guaranteed Funding*) as well as the equilibrium distribution of input levels after allowing for researchers' behavioral responses (*Research Time* and *Total Research Funding*). First focusing on time allocations, the model suggests an optimal distribution of duties and research time that is more unequal than what we actually observe. Conversely, the model suggests that, if anything, the distribution of research funding should be more equal. Appendix Figure C6, which shows the Lorenz curves for the distribution of these inputs, provides another way for illustrating the model's pursuit of a more unequal time distribution and a more equal funding distribution. Appendix Figure C7 shows the distribution of the input wedges across all researchers.

A key limitation of the model is the relatively small degree of heterogeneity allowed for in the consumption (e.g., utility from salary, disutility from work) and demand components (e.g., utility from output) relative to the heterogeneity in the production function. The consumption and demand components are allowed to depend on the PCA-based researcher type index we construct from the survey data. However, this index is only a one-dimensional variable that compresses a lot of heterogeneity in researchers' observable features. Thus, some of the input wedge may not reflect true inefficiency, but instead may actually reflect mis-specification of heterogeneity in the model.

To explore how much such mis-specification may be occurring, Table 5 reports the results from regressions of the input wedge (i.e., actual minus optimal input levels) on different subsets of the vector of observable features. As a simple measure of the variation explained by these observables, the Table reports the R^2 statistics from these regressions. At most, the full vector of observable features can explain roughly 25% of the wedge in time allocations and 12% of the wedge in funding. Thus, a non-trivial portion of the inefficiencies we identify may plausibly be due to mis-specification of heterogeneity along these observable dimensions. Still, the results reported in Table 5 indicate that the majority of input wedges cannot be explained by traditional features of researchers. This suggests that predicting which researchers are over- or under-resourced based only on their observables may be a challenging exercise.

6.4 Case Studies of Potential Frictions

Here, we explore a few specific features of researchers (X_i) that prior work has hypothesized and/or shown to be predictive of frictions in the allocation of inputs to researchers. To do so, we estimate regressions of the form:

$$\text{Actual Input Level}_i = \beta_0 + \beta_1 \text{Optimal Input Level}_i + \beta_2 X_i + \epsilon_i \ , \tag{12}$$

where a perfectly efficient allocation would yield an estimate $\widehat{\beta_1} = 1$ (with no standard error) since actual levels would equal optimal levels, and no other feature would have predictive power (i.e., $\widehat{\beta_2} = 0$). If allocations are not efficient, then $\widehat{\beta_1} < 1$, and if the feature X_i is predictive of misallocation, then $\widehat{\beta_2} \neq 0$. If the feature predicts that the researcher is under-resourced, then $\widehat{\beta_2} < 0$ and vice versa.

First, we focus on a highly scrutinized feature: gender (e.g., Bornmann et al. 2007; European Commission 2015; Van der Lee and Ellemers 2015; Witteman et al. 2019). A large body of work has documented a wide range of biases and frictions facing female researchers when it comes to the acquisition of inputs for their science or the quality of the scientific outputs they have produced. However, while the vast majority of this work rejects null hypotheses of no gender differences and finds female researchers are likely under-resourced, they typically cannot formally quantify how much female researchers are under-resourced. Our model and approach allows us to do just that.

Table 6 reports the results from estimating regressions of the form shown in Equation 12 where X_i is a female gender indicator and we also possibly include all other observable covariates as controls. First, as a benchmark, Column (1) reports the results from a regression without the female indicator or any controls to show how well researchers' actual input levels is predicted by their optimal input levels as implied by the model and our preferred objective. Columns (2) and (3) introduce the female indicator, and whether or not additional controls are included, we estimate a statistically significant negative association whereby female researchers spend approximately 1.3 fewer hours

on their research time and have roughly $14,000 less in total research funding. If these gaps are taken to be fully due to misallocation, it suggests that female researchers are under-resourced on the scale of about 10% (relative to the sample means) compared to their male counterparts on average.

Next, we focus on another question that has received much attention by meta-science scholars: whether the use of grant dollars and publication outcomes as productivity proxies distorts the allocation of inputs (e.g., Lee et al. 2013; Penner et al. 2013; Gralka et al. 2019). Again, we run regressions of the form shown in Equation 12, now where X_i is a vector variables describing researchers' recent grant funding and publication or citation output. As reported in Appendix Table C4, we find some evidence of statistically significant distortions whereby a one standard deviation increase in these traditional proxies leads to an over-resourcing on the scale of roughly 0.1–0.2 standard deviations (approximately 5–10% relative to the sample means).

Thus, we are able to identify statistically significant predictors of the input wedges in these two cases. Researchers are more likely to be under-resourced if they are female, and they are more likely to be over-resourced if their observable input and output measures are higher. However, the R^2 statistics shown in Table 5, Table 6 and Appendix Table C4 convey a seemingly novel point—these observable features, which have received so much attention thus far, explain a very small amount of the misallocation implied by the model. So if the planner is not able to use a methodology like the one we have used here to identify productivity (beliefs) directly, it appears difficult to use standard observable features to predict which researchers are over or under-resourced.

7 Discussion

Our methodology allows us to estimate researchers' productivity beliefs without observing output. With these estimates, we can provide a novel view of allocative efficiency in science. While we do find current input allocations to be positively correlated with allocations that maximize plausible objectives, we also find that there are large potential gains from more efficiently allocating inputs. In the counterfactuals we explore, total scientific output per-research-hour or per-research-dollar could be more than doubled. The limitations of our methodology motivate us to interpret these magnitudes as plausible upper bounds on the gains from more efficient of scientific inputs. More importantly, these large potential gains provide a glimmer of hope amid trends of declining R&D productivity across most sectors of the economy (Bloom et al. 2020) and the persistently growing burden of knowledge that raises the cost of conducting frontier science (Jones 2009).

To our knowledge, our analyses are some of the first comprehensive evaluations of productivity and allocative efficiency in science. Our approach is focused on researchers' *beliefs* about their own productivity and therefore their *beliefs* about how well inputs are allocated. On one hand this is a crucial limitation. However, science is inherently about production on the frontier, and, therefore, the optimal mechanisms for identifying productive researchers and allocating them more inputs will

need to tackle this notion of researchers' forecasting their productivity. We believe the methods we implement here are an important step in that direction.

Interestingly, we find that the gains from an unconstrained total research budget, where the limits on aggregate output are constrained only by researchers' time, are quite small once the existing budget has been efficiently allocated. This could have important policy implications—instead of debating how much larger total investments in science should be, our results suggest that much larger gains could be had from more accurately identifying productive researchers and allocating them more inputs. However, we still lack a good understanding of the costs of these efforts. In our data, it appears quite difficult to identify under- and over-resourced researchers using the observable features traditionally available. Our methodology provides a new way of identifying researchers' productivity beliefs, which may prove useful in these reallocation efforts.

Our approach to identifying productivity here has deep roots in the fields of marketing (i.e., using surveys to identify demand functions), labor (i.e., surveys of time use), industrial organization (i.e., production function estimation), and macroeconomics (i.e., models of factor misallocation). When combined, this new approach allow us to overcome many of the challenges that have long plagued our understanding of productivity and efficiency in science. However, this approach introduces new assumptions and challenges to grapple with. Still, we hope this method is viewed as a complementary approach to understanding productivity in science.

One of the most important next steps in this line of work is likely to be developing methods to ensure that the agents being evaluated using this methodology report their willingness to pay truthfully. Of course, the theoretical underpinnings of how to elicit true willingness-to-pay estimates have long been established (Becker et al. 1964). However, the magnitudes involved in the scenarios here involve trade-offs on the scale of tens or hundreds of thousands of dollars. Thus, developing the practical details of incentivizing truthful responses from producers at this scale would be a fruitful endeavor.

More broadly, our approach may prove useful in other settings that face similar challenges in estimating productivity and efficiency. There are many markets populated by a large number of producers acquiring inputs in a highly decentralized way to produce outputs that are not easy to observe. For example: in developing economies, accurate producer-level data on outputs can be difficult to obtain (e.g., Tybout 2000); in the realm of entrepreneurship, many organizations never produce any observable output before they exit (e.g., Decker et al. 2014); and in non-profit sectors, the output is often such a high-dimensional object that achieving consensus on a suitable proxy is challenging (e.g., Philipson and Lakdawalla 2001). In all of these examples, our methodology could provide a way forward to better understanding the distribution and determinants of productivity.

References

Adams, J. D. (1990). Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy*, 98(4):673–702.

Adams, J. D. and Griliches, Z. (1998). Research productivity in a system of universities. *Annals of Economics and Statistics*, 49-50:127–162.

Andrews, M. J. (2023). How do institutions of higher education affect local invention? Evidence from the establishment of US colleges. *American Economic Journal: Economic Policy*, 15(2):1–41.

Arora, A., Belenzon, S., Patacconi, A., and Suh, J. (2020). The changing structure of American innovation: Some cautionary remarks for economic growth. *Innovation Policy and the Economy*, 20(1):39–93.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pages 609–626. Princeton University Press.

Azoulay, P., Fons-Rosen, C., and Graff Zivin, J. S. (2019a). Does science advance one funeral at a time? *American Economic Review*, 109(8):2889–2920.

Azoulay, P., Graff Zivin, J. S., Li, D., and Sampat, B. (2019b). Public R&D investments and private-sector patenting: Evidence from NIH funding rules. *Review of Economic Studies*, 86(1):117–152.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–1144.

Bornmann, L., Mutz, R., and Daniel, H.-D. (2007). Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1(3):226–238.

Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24.

Digital Science (2018). Dimensions [software]. Available from https://app.dimensions.ai. Accessed in 2023, under licence agreement.

Dizon-Ross, R. and Jayachandran, S. (2022). Improving willingness-to-pay elicitation by including a benchmark good. *AEA Papers and Proceedings*, 112:551–555.

European Commission (2015). She figures. Technical report, Directorate-General for Research and Innovation, Luxembourg: Publications Office.

Gralka, S., Wohlrabe, K., and Bornmann, L. (2019). How to measure research efficiency in higher education? research grants vs. publication output. *Journal of Higher Education Policy and Management*, 41(3):322–341.

Hausman, N. (2022). University innovation and local economic growth. *Review of Economics and Statistics*, 104(4):718–735.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics*, 6(1):735–770.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4):1403–1448.

Jaffe, A. B. (1989). Real effects of academic research. *American Economic Review*, 79(5):957–970.

Jaffe, A. B. and Lerner, J. (2001). Reinventing public R&D: Patent policy and the commercialization of national laboratory technologies. *RAND Journal of Economics*, pages 167–198.

Jena, A. B. and Philipson, T. J. (2008). Cost-effectiveness analysis and innovation. *Journal of Health Economics*, 27(5):1224–1236.

Jones, B. F. (2009). The burden of knowledge and the "Death of the renaissance man": Is innovation getting harder? *Review of Economic Studies*, 76(1):283–317.

Lakdawalla, D. N., Sun, E. C., Jena, A. B., Reyes, C. M., Goldman, D. P., and Philipson, T. J. (2010). An economic evaluation of the war on cancer. *Journal of Health Economics*, 29(3):333–346.

Lee, C. J., Sugimoto, C. R., Zhang, G., and Cronin, B. (2013). Bias in peer review. *Journal of the American Society for information Science and Technology*, 64(1):2–17.

Mansfield, E. (1998). Academic research and industrial innovation: An update of empirical findings. *Research Policy*, 26(7-8):773–776.

Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. The University of Chicago Press.

Myers, K. R., Tham, W. Y., Cohodes, N., Lakhani, K., Mural, R., Thursby, J. G., Thursby, M. C., and Xu, Y. (2023). New facts and data about US research professors. *Mimeo*.

Myers, K. R., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M. C., Schiffer, P., Walsh, J. T., Lakhani, K. R., and Wang, D. (2020). Unequal effects of the COVID-19 pandemic on scientists. *Nature Human Behaviour*, 4(9):880–883.

National Science Foundation (2023). Higher Education Research and Development Survey (HERD). Available from https://www.nsf.gov/statistics/srvyherd/. Accessed in 2023.

Nelson, R. (1959). The simple economics of basic scientific research. *Journal of Political Economy*, 67(3):297–306.

Nordhaus, W. D. (2004). Schumpeterian profits in the american economy: Theory and measurement.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263.

Penner, O., Pan, R. K., Petersen, A. M., Kaski, K., and Fortunato, S. (2013). On the predictability of future impact in science. *Scientific reports*, 3(1):3052.

Philipson, T. and Jena, A. (2006). Surplus appropriation from R&D and health care technology assessment procedures.

Philipson, T. and Lakdawalla, D. (2001). Medical care output and productivity in the nonprofit sector. In Cutler, D. M. and Berndt, E. R., editors, *Medical care output and productivity*, pages 119–140. University of Chicago Press.

Popp, D. (2017). From science to technology: The value of knowledge from different energy research institutions. *Research Policy*, 46(9):1580–1594.

Qiu, Y. J. (2023). The Matthew effect, research productivity, and the dynamic allocation of NIH grants. *RAND Journal of Economics*, 54(1):135–164.

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10.

Rosenbloom, J. L., Ginther, D. K., Juhl, T., and Heppert, J. A. (2015). The effects of Research & Development funding on scientific productivity: Academic chemistry, 1990-2009. *PLoS ONE*, 10(9):e0138176.

Sarsons, H. and Xu, G. (2021). Confidence men? Evidence on confidence and gender among top economists. *AEA Papers and Proceedings*, 111:65–68.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9):628–638.

Shapin, S. (1995). Here and everywhere: Sociology of scientific knowledge. *Annual Review of Sociology*, 21(1):289–321.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature*, 49(2):326–365.

Tybout, J. R. (2000). Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic Literature*, 38(1):11–44.

Valero, A. and Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, 68:53–67.

Van der Lee, R. and Ellemers, N. (2015). Gender contributes to personal research funding success in the netherlands. *Proceedings of the National Academy of Sciences*, 112(40):12349–12353.

Wang, E. C. and Huang, W. (2007). Relative efficiency of R&D activities: A cross-country study accounting for environmental factors in the DEA approach. *Research Policy*, 36(2):260–273.

Witteman, H. O., Hendricks, M., Straus, S., and Tannenbaum, C. (2019). Are gender gaps due to evaluations of the applicant or the science? a natural experiment at a national funding agency. *The Lancet*, 393(10171):531–540.

Zuckerman, H. (1988). The sociology of science. In Smelser, N., editor, *Handbook of Sociology*, pages 511–574. Sage Publications.

Tables & Figures

Table 1: Summary Statistics—Salary, Funding, Time, and Fields

	mean	*s.d.*	*p*50
Salary and research funds, $/year			
Total salary	159,028.23	74,516.13	140,000.00
Guaranteed & existing funding	52,223.08	83,420.47	5,000.00
Fundraising expectations	93,909.19	144,013.69	20,000.00
Work time, hrs./week			
Total work	48.55	10.00	48.00
Research	18.51	9.60	17.40
Fundraising	4.47	5.09	2.65
Administration	7.48	6.17	5.80
Teaching and other work	18.09	9.78	17.20
Major field, {0,1}			
Engineering, math, & related	0.17	0.37	
Humanities & related	0.19	0.39	
Medical & health sciences	0.28	0.45	
Natural sciences	0.16	0.37	
Social sciences	0.21	0.40	

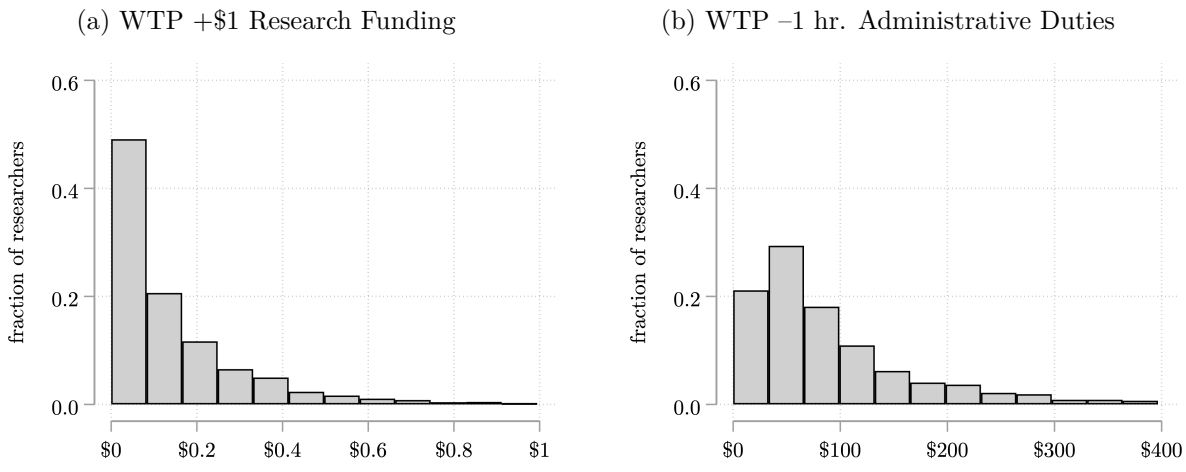
Note: Reports summary statistics for 4,003 researcher-level observations. Unless otherwise noted, all variables are continuous and bound below by zero. *{0,1}* indicates binary variables.

Table 2: Summary Statistics—Subjective Output

	mean	*s.d.*
Intended research outputs, {0,1,2}		
Journal articles	1.87	0.37
Books	0.52	0.68
Materials or methods	0.70	0.68
Products	0.47	0.64
Intended research audience, {0,1,2}		
Academic peers	1.88	0.36
Policymakers	0.84	0.69
Businesses	0.54	0.62
General public	0.83	0.62
Riskiness of own research, [0,10]		
Own belief	4.63	2.34
Belief of peers' beliefs	4.57	2.37
Theoretical vs. empirical, [0,10]		
Ask [0] or answer [10] questions	4.89	2.54

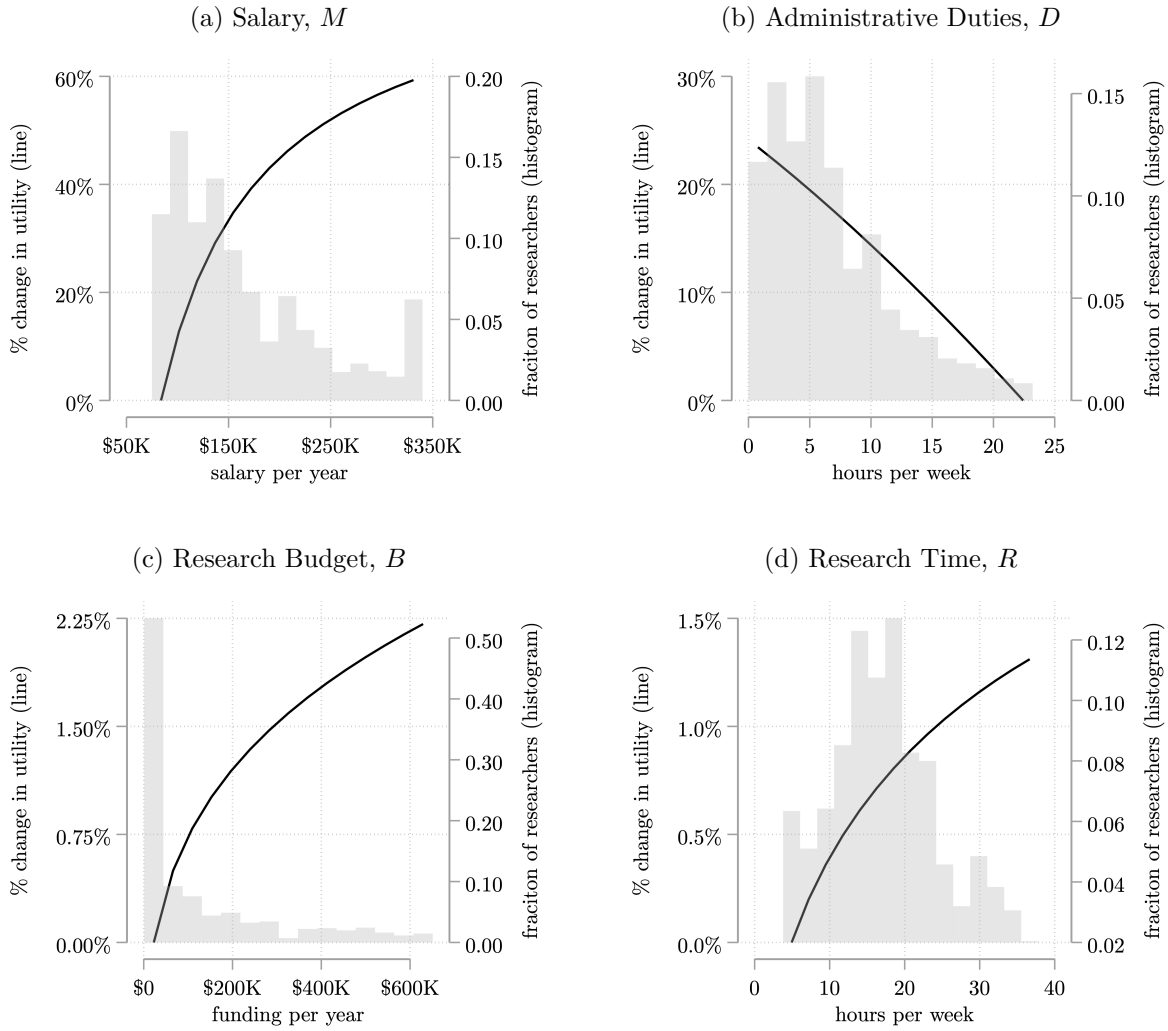
Note: Reports summary statistics for 4,003 researcher-level observations. The intended research outputs and audience variables are coded responses to questions of the form *How often are the following the intended output / audience of your research*, where responses are coded as follows: *Rarely*=0, *Sometimes*=1, *Very often*=2. The question about risk used a scale where 0 indicated no risk and 10 indicated very high risk. The question about theory versus empirics used a scale where 0 indicated that the researcher focused on asking new questions and 10 indicated that the researcher focsed on answering existing questions.

Figure 1: Willingness to Pay Experiment Responses



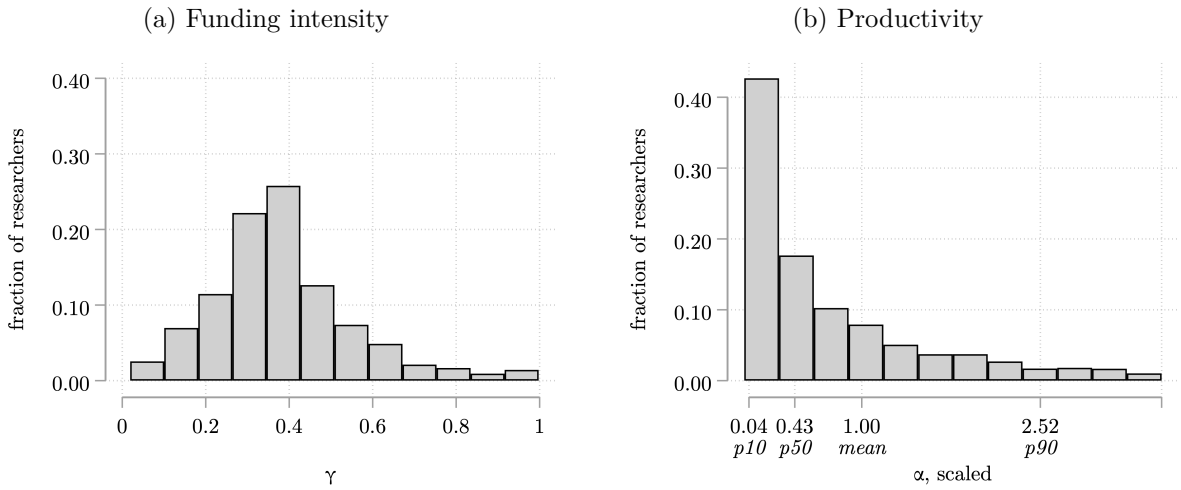
Note: Shows the distribution of researchers' stated WTP for $1 more of additional research funds (Panel a) and 1 hour less of administrative duties (Panel b). Approximately 1% of the upper tails have been trimmed for visibility.

Figure 2: Utility Function Visualization



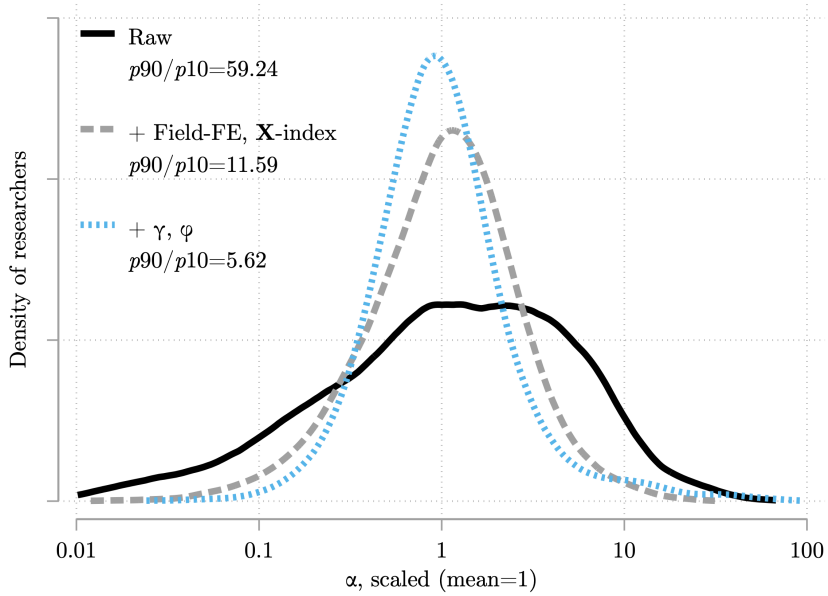
Note: Shows the percent change in utility in four of the key model variables (Panels a–d) holding all other variables and parameters fixed at the sample means. The black line (left y axis) shows the percent change in utility as the focal variable increases from the 10th percentile to the 90th percentile. The histogram (right y axis) shows the distribution of the focal variable in the sample. Note the different scales of all y axes.

Figure 3: Production Function Estimates



Note: Shows the distribution of researcher-specific estimates of the production function parameters γ_i (funding intensity; Panel a) and α_i (productivity; Panel b). In Panel (a), larger values indicate more funding intensity and $\gamma \in [0, 1]$. In Panel (b), the top 5% of productivities have been trimmed for visibility and estimates are scaled into units where the mean is normalized to equal 1 (with the 10th, 50th, and 90th percentiles of the distribution noted on the x axis).

Figure 4: Conditional Productivity Distributions



Note: Shows the distribution of researchers' productivity where units have been rescaled so that the mean equals 1; note the log scale. In addition to (*i*) the distribution of the raw values, productivities are also shown (*ii*) conditional on major field fixed effects and the type index, and (*ii*) conditional on funding intensity, fundraising efficiency, major field fixed effects and the type index. 90-10 percentile ratios (of the non-log transformed) productivities are also shown.

Table 3: Production with Actual and Optimized Allocations

	Current allocation	Optimized allocations					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Research inputs							
Research hrs./week, avg.	18.5	–11%	+18%	+14%	+27%	+25%	+16%
Research hrs./week, s.d.	9.6	+6.0%	+18%	+43%	+43%	+39%	+41%
Budget $-K/year, avg.	147.1	0%	0%	0%	0%	0%	+14%
Budget $-K/year, s.d.	206.9	+14%	–24%	–20%	–18%	–20%	–3.9%
Research output							
Output, avg.	*n.r.*	+1.3%	+150%	+160%	+160%	+160%	+160%
Output, s.d.	*n.r.*	+0.7%	+43%	+44%	+44%	+44%	+44%
Output per hr.	*n.r.*	+14%	+110%	+130%	+100%	+110%	+120%
Output per $	*n.r.*	+1.4%	+150%	+160%	+160%	+160%	+130%
Value							
Researcher utility, avg.	*n.r.*	+3.6%	+1.0%	+4.7%	+4.5%	+2.1%	+4.7%
Researcher utility, s.d.	*n.r.*	–5.7%	+0.3%	–5.4%	–5.2%	–2.4%	–5.4%
Researcher utility per hr.	*n.r.*	–8.3%	+16%	+16%	+25%	+21%	+18%
Researcher utility per $	*n.r.*	+3.6%	+1.0%	+4.8%	+4.5%	+2.3%	+16%
Social value, avg.	*n.r.*	+3.4%	+12%	+16%	+16%	+13%	+16%
Social value, s.d.	*n.r.*	–4.0%	+23%	+20%	+21%	+21%	+20%
Social value per hr.	*n.r.*	–8.5%	+25%	+26%	+34%	+30%	+28%
Social value per $	*n.r.*	+3.4%	+12%	+16%	+16%	+13%	+26%
Input reallocation							
Research hrs./week		82%	62%	47%	43%	52%	51%
Budget $-K/year		72%	46%	44%	50%	49%	50%
Objective, max		U	U	U	W	Y	U
Reallocate D		✓		✓	✓	✓	✓
Reallocate G			✓	✓	✓	✓	✓
Unconstrained B							✓

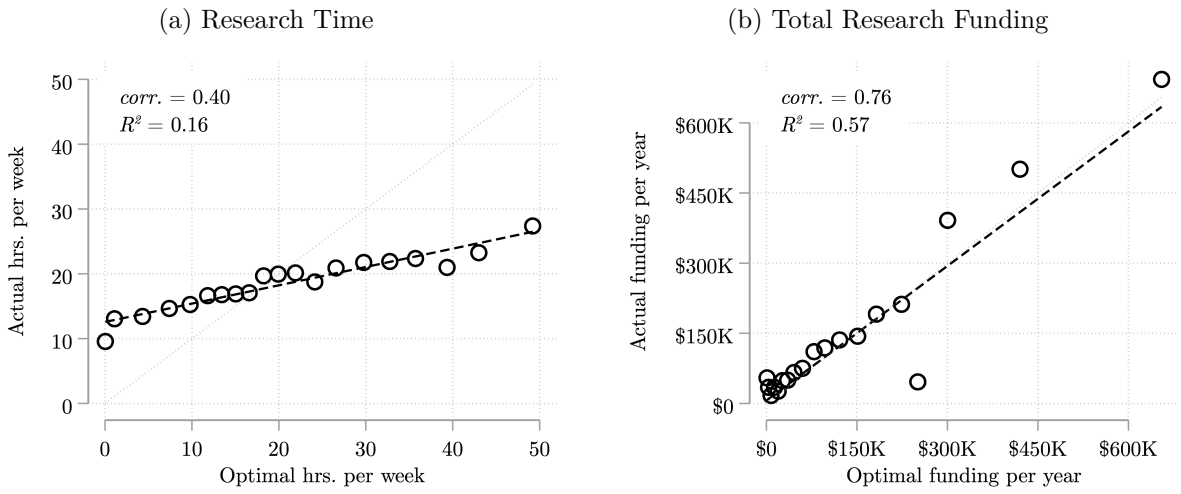
Note: Reports summary statistics for inputs under current allocations (Col. 1). The first three sets of rows in Columns 2–7 report the percentage change in research inputs (*Research inputs*), outputs (*Research outputs*), and utility (*Value*) under alternative allocations. The *Input Reallocation* rows report the amount of inputs reallocated expressed as a percentage of the total level of the input (e.g., 50% implies that half of all dollars are moved from one researcher to another). The bottom sets of rows outline the objective and constraints of the six different counterfactual allocations explored in Columns 2–7. The three different objectives explored are maximizing researchers' private utility (U), social welfare (W), or output (Y). D refers to administrative duties, and G refers to guaranteed research funding. *Unconstrained* B indicates the scenario when the total research budget is left unconstrained and so the total amount of funding in the market is limited only by researchers' fundraising choices. All optimized allocations allow for researchers' behavioral responses after D and/or G have been reallocated.

Table 4: Growth in Funding with Current Allocations Needed to Produce the Same Output as Optimized Allocation of Current Funding

	Mechanical	Behavioral
Δ *Annual guaranteed funding, G*		
Sample average	+$160K (109%)	+$155K (105%)
Δ *Annual total funding, B*		
Sample average	+$160K (42%)	+$155K (41%)
Sample total	+$249M	+$241M
Population total	+$14.9B	+$14.4B

Note: Reports the increase in guaranteed (G) and total (B) research funding under current allocations necessary to achieve the same growth in output achieved by the reallocation of current input levels that maximizes researchers' utility (+160%). Sample averages on a per-researcher basis are reported in addition to sample totals (summing over all in-sample researchers) and implied population totals (scaling the sample up to the population size per the survey response rate). The *Mechanical* scenario does not allow researchers to re-optimize their time allocations and the *Behavioral* scenario does.

Figure 5: Actual Versus Optimal Input Correlations



Note: Shows the binned scatterplot of actual and optimal input levels also reporting the pairwise correlation between the two variables and the R^2 statistic from a regression of the actual level on the optimal level. Based on the counterfactual described in Column (4) of Table 3.

Table 5: Proportion of Input Wedge Variation Correlated with Observables

	Research time				Research budget			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R^2	0.03	0.06	0.18	0.25	0.01	0.06	0.06	0.12
X: position	✓			✓	✓			✓
X: output		✓		✓		✓		✓
X: socio-demog.			✓	✓			✓	✓
N obs.	4,003	4,003	4,003	4,003	4,003	4,003	4,003	4,003

Note: Reports R^2 statistics from regressions of the input wedge (i.e., actual minus optimal input levels) on different subsets of covariates **X** used to construct the type index; *position* covariates include researchers' rank and tenure status; *output* covariates include researchers' Likert-scale based reporting of the intended outputs (e.g., publications, books) and audiences (e.g., peers, policymakers) of their research; *socio-demog.* covariates include measures of household structure, race/ethnicity, nationality, and gender.

Table 6: Input Wedges and Researcher Gender

	(1)	(2)	(3)
Panel (a): Actual research time			
Optimal level	0.296***	0.296***	0.327***
	(0.0105)	(0.0105)	(0.0111)
Female		–1.294***	–0.993***
		(0.287)	(0.284)
X			✓
R^2	0.18	0.18	0.30
N obs.	3,800	3,800	3,800
Panel (b): Actual research funding			
Optimal level	0.950***	0.947***	0.884***
	(0.0100)	(0.0101)	(0.0127)
Female		–13695.1***	–11009.9***
		(4307.9)	(4266.4)
X			✓
R^2	0.60	0.60	0.64
N obs.	3,800	3,800	3,800

Note: Reports results from regressions of actual input levels on optimal input levels and a binary indicator for female gender. Research time is in units of hours per week, and research funding is in units of $ per year. Robust standard errors reported; $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

A Appendix: Additional Survey Statistics and Comparisons

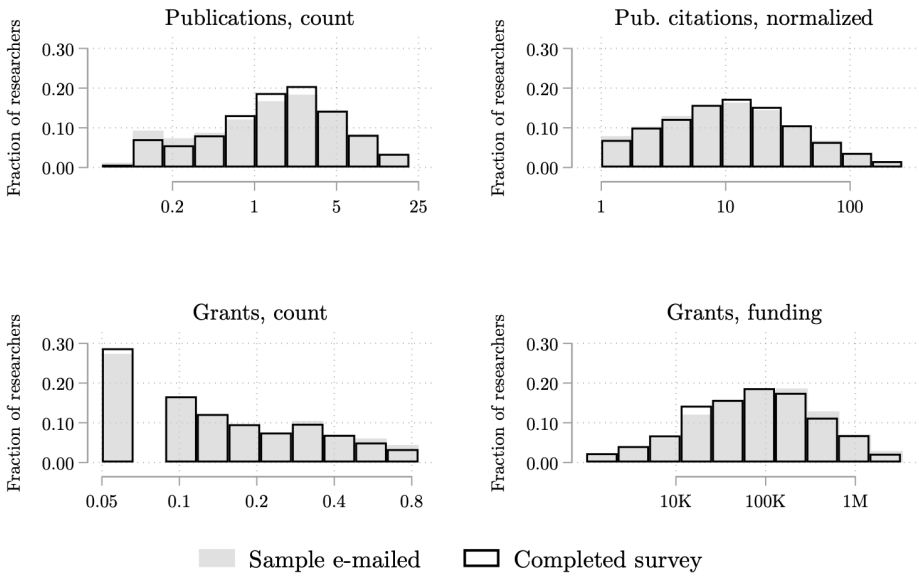
Table A1: Survey Completion per Randomized Treatments

	(1)
Either incentive	0.00297**
	(0.00117)
Both incentives	0.00797***
	(0.00141)
1 reminder	0.0112***
	(0.00114)
2 reminders	0.0188***
	(0.00119)
Constant	0.0197***
	(0.00107)
N obs.	131,672

Note: Reports estimates from a regression of a binary indicator of survey completion on binary indicators for the randomized incentives and reminders including observations for all researchers emailed. Robust standard errors reported; $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Figure A1: Sample Representativeness per Publication and Grant Funding Measures

(a) Distribution Comparisons



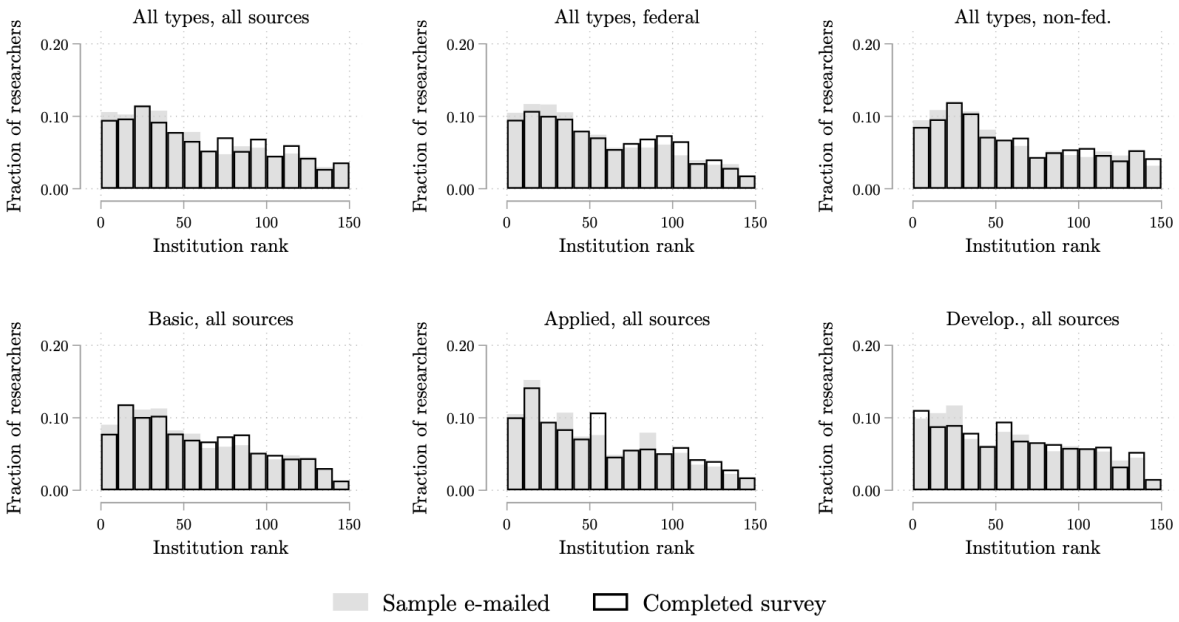
(b) Regression Comparisons

	Pub., count	Pub. cites, normalized	Grant, count	Grant, total $
	(1)	(2)	(3)	(4)
Completed survey	–0.111	–0.575	0.00878**	–5568.8
	(0.0688)	(0.807)	(0.00343)	(8519.4)
Constant	2.906***	20.59***	0.117***	121272.4***
	(0.0158)	(0.185)	(0.000715)	(2093.0)
% diff.	–3.8%	–2.8%	7.5%	–4.6%
N obs.	87,000	87,000	87,000	87,000

Note: Panel (a) shows the distribution of publication and grant outcomes split by whether the researcher was emailed (i.e., the full sample) versus those who completed the survey; note the log x axes. Panel (b) reports estimates from a regression of each measure on a binary indicator for whether the researcher completed the survey; *% diff.* reports this difference in percentage terms. Robust standard errors reported; $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Figure A2: Sample Representativeness per Institutional Funding Measures

(a) Distribution Comparisons



(b) Regression Comparisons

	All sources, by type			All types, by source		
	All	Federal	Non–fed.	Basic	Applied	Develop.
	(1)	(2)	(3)	(4)	(5)	(6)
Completed survey	–35.01***	–21.30***	–13.71***	–24.31***	–7.170***	–3.812***
	(6.953)	(4.511)	(3.062)	(4.867)	(2.492)	(1.328)
Constant	649.1***	351.7***	297.4***	413.2***	176.7***	60.46***
	(1.308)	(0.863)	(0.578)	(0.917)	(0.459)	(0.254)
% diff.	–5.4%	–6.1%	–4.6%	–5.9%	–4.1%	–6.3%
N obs.	130,785	130,785	130,785	130,785	130,735	128,169

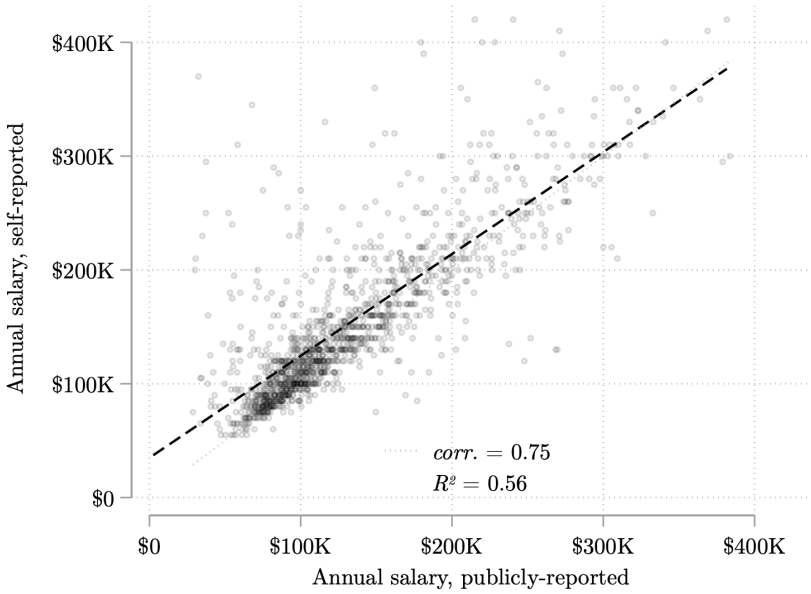
Note: Panel (a) shows the distribution of institutional funding measures split by whether the researcher was emailed (i.e., the full sample) versus those who completed the survey; note the log x axes. Panel (b) reports estimates from a regression of each measure on a binary indicator for whether the researcher completed the survey; *% diff.* reports this difference in percentage terms. Robust standard errors reported; $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table A2: Distribution of Minor Fields of Study

	share
Agriculture	0.02
Biology	0.05
Business	0.05
Chemistry	0.02
Communication	0.03
Computer science	0.02
Economics	0.02
Education	0.04
Engineering	0.06
Geography	0.05
Humanities	0.11
Law	0.02
Mathematics	0.03
Medical school	0.21
Medicine and health	0.07
Other social sciences	0.04
Physics	0.04
Political science	0.04
Psychology	0.04
Sociology	0.04

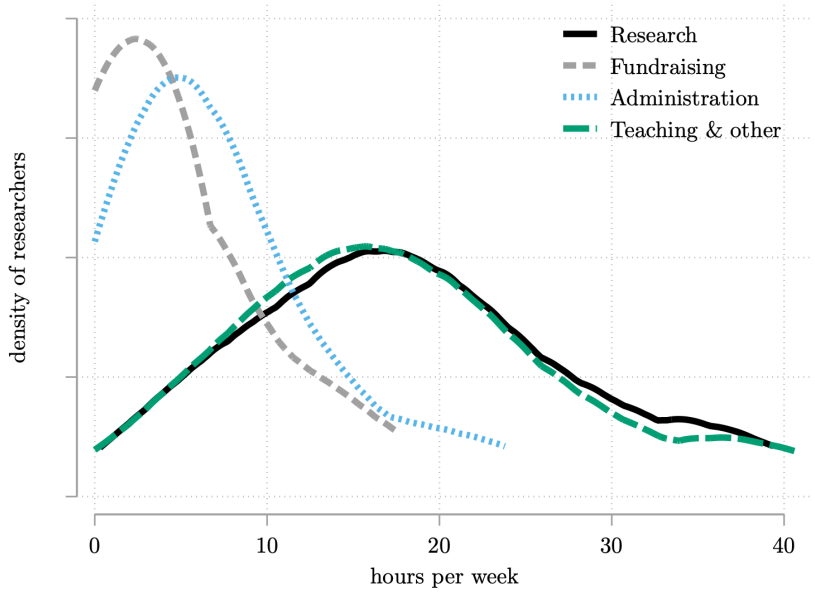
Note: Reports the share of 4,003 researcher-level observations in each minor field.

Figure A3: Correlation of Self- and Publicly-reported Annual Salaries



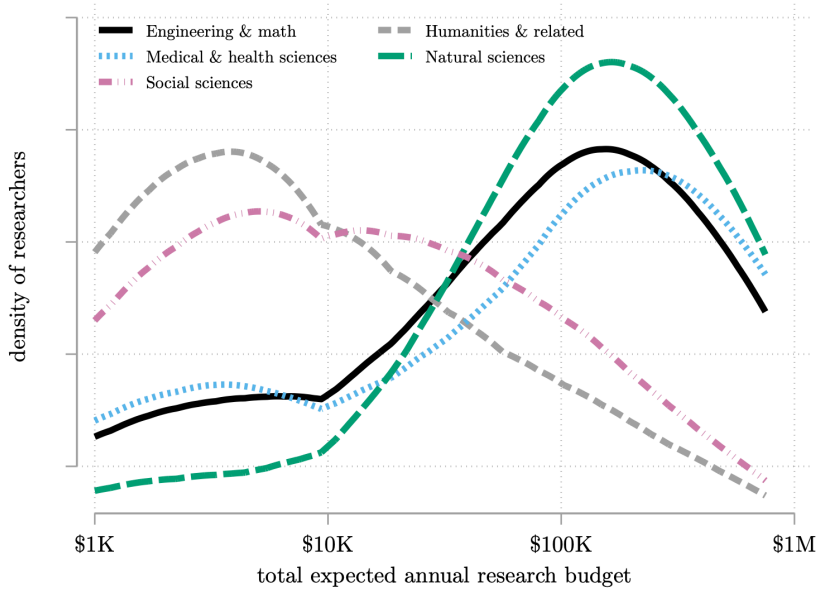
Note: Shows the publicly- and self-reported annual salaries for 1,369 in-sample researchers whose salaries were located in public reportings, and reports the pairwise correlation alongside the R^2 statistic from a regression of self-reported on publicly-reported salary.

Figure A4: Time Allocation Distributions



Note: Shows the distribution of researchers' hours per week spent on four categories of work tasks.

Figure A5: Research Funding Distributions by Major Fields of Study



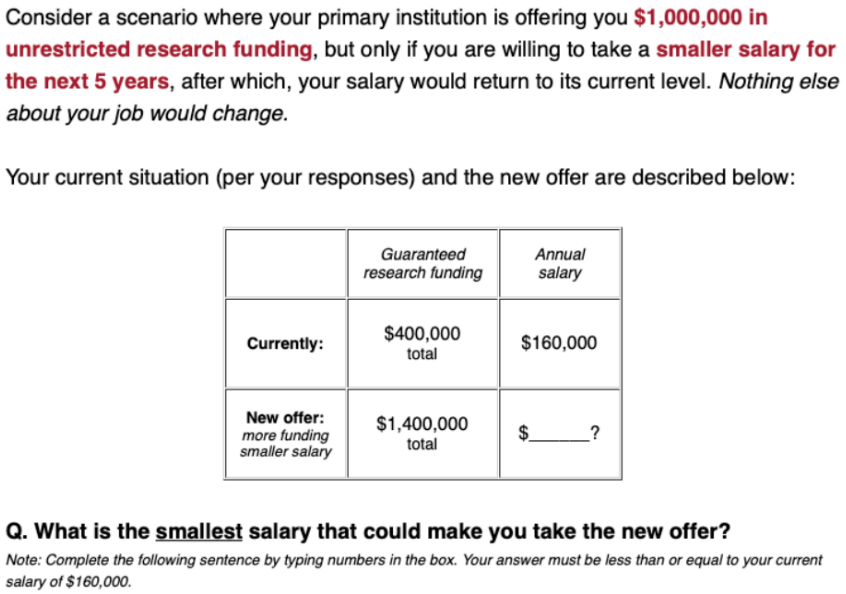
Note: Shows the distribution of researchers' total expected annual research budget (including guaranteed funds plus fundraising expectations) split by the researchers' major field of study; note the log scale.

Table A3: Summary Statistics of Other Variables

	mean	*s.d.*
From HERD: Institution-level R&D		
Total R&D, $M	611.32	451.18
R&D per researcher, $M	0.61	0.86
Share federal gov.'t R&D, [0,1]	0.52	0.12
Share basic R&D, [0,1]	0.63	0.20
From Dimensions: Research output		
Publications per year	5.28	7.09
Citations per year	23.07	49.10
Co-authors per publication per year	9.09	68.40
Position details		
Assistant professor, {0,1}	0.26	0.44
Associate professor, {0,1}	0.26	0.44
Full professor, {0,1}	0.41	0.49
Other rank, {0,1}	0.07	0.25
Not on tenure track, {0,1}	0.19	0.39
Pre-tenure, {0,1}	0.22	0.42
Tenured, {0,1}	0.58	0.49
Years until next contract eval.	3.68	1.82
Duration of contract	2.45	1.93
Gender identity, {0,1}		
Female	0.40	0.49
Male	0.55	0.50
Other or N.R	0.05	0.22
Racial/ethnic identity, {0,1}		
Asian	0.13	0.33
Black	0.03	0.18
Hispanic	0.06	0.23
White	0.77	0.42
Other or N.R	0.05	0.21
Citizenship, {0,1}		
Citizen, domestic-born	0.71	0.45
Citizen or perm resident, foreign-born	0.24	0.43
Other or N.R citizenship	0.05	0.21
1st–3rd generation in U.S.	0.30	0.46
Other or N.R generation in U.S.	0.70	0.46
Other covariates		
Age	48.78	12.05
Household total income	260,336.00	215,388.31
Married or domestic partnership, {0,1}	0.82	0.39
Single, {0,1}	0.13	0.34
Other or N.R relationship, {0,1}	0.05	0.22
Dependents in household	0.98	1.13
Risk-taking in personal life, [0,10]	5.26	2.13

Note: Reports summary statistics for 4,003 researcher-level observations. *From HERD* indicates variables from National Science Foundation (2023). *From Dimensions* indicates variables from the Digital Science (2018) dataset. *N.R.* stands for Not Reported.

Figure A6: Example Image of Survey Experiment



Note: Shows a screenshot of the survey experiment designed to solicit researchers' willingness to trade off their salary for additional guaranteed funding.

Figure A7: Correlations of Implied Valuations



Note: Shows a binned scatterplot and line-of-fit for the relationship between researchers' implied hourly wage and their willingness to pay for 1 less hour of administrative duties (Panel a), and the relationship between how many hours per week a researcher spends on fundraising and their willingness to pay for $1 more of additional research funds conditional on their implied hourly wage (Panel b).

Table A4: Proportion of Response Variation Correlated with Model and Non-model Variables

	WTP +$250K		WTP +$1M		WTP +admin.		WTP –admin.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R^2	0.95	0.96	0.87	0.91	0.82	0.88	0.96	0.97
Model vars.	✓	✓	✓	✓	✓	✓	✓	✓
X		✓		✓		✓		✓
N obs.	4,003	4,003	4,003	4,003	4,003	4,003	4,003	4,003

Note: Reports the R^2 statistics from regressions of researchers' responses to the four experiments (i.e., their willingness to pay for the alternative scenarios) on different combinations of variables: *Model vars.* includes the state and choice variables of the model; **X** includes the full vector of variables that comprise the type index.

Table A5: Potential Role of Non-response Bias and Stated Preference WTP Bias

	WTP +$250K		WTP +$1M		WTP +admin.		WTP –admin.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inverse Mills Ratio, std.		–0.000858		–0.00391		–0.00466		–0.000835
		(0.00379)		(0.00593)		(0.00684)		(0.00327)
Benchmark WTP, std.		–0.0126***		–0.0258***		0.0134*		–0.0102***
		(0.00411)		(0.00644)		(0.00742)		(0.00355)
R^2	0.95	0.95	0.87	0.87	0.82	0.82	0.96	0.96
Model vars.	✓	✓	✓	✓	✓	✓	✓	✓
N obs.	4,003	4,003	4,003	4,003	4,003	4,003	4,003	4,003

Note: Reports the estimates from regressions of researchers' responses to the four experiments (i.e., their willingness to pay for the alternative scenarios) on a control for non-response bias in the form of the Inverse Mills Ratio and a control for bias in individuals' stated WTP in the form of their WTP for the benchmark good (i.e., high-speed internet at home); all variables are standardized. *Model vars.* includes the state and choice variables of the model. Robust standard errors reported; $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$.

B Appendix: Model Details

B.1 Derivation of Policy Functions

The policy functions $\mathcal{R}(\mathbf{S}_i, \boldsymbol{\theta}_i, \boldsymbol{\mu}_i)$, $\mathcal{F}(\mathbf{S}_i, \boldsymbol{\theta}_i, \boldsymbol{\mu}_i)$ and $\mathcal{H}(\mathbf{S}_i, \boldsymbol{\theta}_i, \boldsymbol{\mu}_i)$ characterize the solution (R_i^*, F_i^*, H_i^*) to problem (1) for each individual scientist i as a function of states $\mathbf{S}_i$, attributes $\boldsymbol{\theta}_i$, and parameters $\boldsymbol{\mu}_i$. Moreover, they determine indirect utility $V_i^* = \mathcal{V}(\mathbf{S}_i, \boldsymbol{\theta}_i, \boldsymbol{\mu}_i)$ at current allocations.

Once substituting the constraints, we can express the utility-maximization problem as

$$\mathcal{V}(\cdot,\cdot,\cdot) = \max_{F_i, H_i} U_{1i}(M_i) + U_{2i}(\alpha_i \left(B_{\min} + G_i + \phi_i F_i\right)^{\gamma_i} \left(H_i - F_i - D_i\right)^{1-\gamma_i}) - U_{3i}(H_i, D_i) \tag{B1}$$

with the policy functions $\mathcal{F}(\cdot,\cdot,\cdot)$ and $\mathcal{H}(\cdot,\cdot,\cdot)$ solving the optimality conditions:

$$\frac{\partial U_{2,i}(Y_i)}{\partial Y_i}\frac{\partial Y_i}{\partial F_i} + \lambda_{F,i} = 0 \tag{B2}$$

$$\frac{\partial U_{2,i}(Y_i)}{\partial Y_i}\frac{\partial Y_i}{\partial H_i} - \frac{\partial U_{3,i}(H_i, D_i)}{\partial H_i} - \lambda_{H,i} = 0\,, \tag{B3}$$

and $\mathcal{R}(\cdot,\cdot,\cdot)$ being residually determined based on the time-constraint (1c).

We derive model's policy functions describing how scientists' optimal time allocation vary with state variables and parameters. We start from model's optimality conditions (B3) and (B3) and evaluate them using the chosen functional forms described in the main text. We obtain the conditions:

$$Y_i^{1-\eta_i}\left[\gamma_i\phi_i(B_{\min} + G_i + \phi_i F_i)^{-1} - (1-\gamma_i)(H_i - D_i - F_i)^{-1}\right] + \lambda_{F,i} = 0 \tag{B4}$$

$$Y_i^{1-\eta_i}(H_i - D_i - F_i)^{-1}(1-\gamma_i) - \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0\,. \tag{B5}$$

We characterize the policy functions in two intervals of H_i. The first is $H_i \in (D_i, H_{i,F>0}]$ and the second is $H_i \in (H_{i,F>0}, H_{\max}]$, depending on whether the optimal time allocation features a corner solution for fundraising activity ($F_i = 0$) or an interior one ($F_i > 0$). We identify the threshold $H_{i,F>0}$ below which the optimal fundraising time is zero by solving (B5) for F_i as a function of H_i under the hypothesis that F_i is strictly positive, i.e., $\lambda_{F,i} = 0$:

$$F_i = \gamma_i(H_i - D_i) - \frac{1-\gamma_i}{\phi_i}(B_{\min} + G_i)\,. \tag{B6}$$

Therefore, (B6) implies that fundraising time is strictly positive as long as:

$$H_i > D_i + \frac{1-\gamma_i}{\gamma_i\phi_i}(B_{\min} + G_i)\,, \tag{B7}$$

which identifies the threshold $H_{i,F>0}$, which is strictly larger than D_i as long as $\gamma_i < 1$. As a consequence, we can solve for optimal hours H_i using equation (B5), which takes the piece-wise

functional form:

$$\begin{cases} (1-\gamma_i)[\alpha_i(B_{\min}+G_i)^{\gamma_i}]^{1-\eta_i}(H_i-D_i)^{(1-\gamma_i)(1-\eta_i)-1} - \psi(H_i-D_i+D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0 & \text{if } H_i \leq H_{i,F>0} \\ \left[\alpha_i(\phi_i\gamma_i)^{\gamma_i}(1-\gamma_i)^{1-\gamma_i}\right]^{1-\eta_i}\left(H_i - D_i\frac{B_{\min}+G_i}{\phi_i}\right)^{-\eta_i} - \psi(H_i-D_i+D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0 & \text{if } H_i > H_{i,F>0} \end{cases} \tag{B8}$$

and defines an implicit solution $\mathcal{H}_i$ for H_i as a function of all parameters and state variables. Specifically, (B8) is continuous but not differentiable at $H_{i,F>0}$. Moreover, the Lagrange multiplier $\lambda_{H,i}$ equals zero for $H_i < H_{\max}$. Finally, previous equation implies that H_i is always strictly larger than D_i because $\lim_{H_i \to D_i}$ (B8) $= +\infty$. Hence, the non-negativity constraint on research time is always met.

Given the policy function $\mathcal{H}_i$ for hours, the functions defining optimal allocation to fundraising time and research time can be derived as:

$$\mathcal{F}_i = \begin{cases} 0 & \text{if } \mathcal{H}_i \leq H_{i,F>0} \\ \gamma_i(\mathcal{H}_i - D_i) - \frac{1-\gamma_i}{\phi_i}(B_{\min}+G_i) & \text{if } \mathcal{H}_i > H_{i,F>0} \end{cases} \tag{B9}$$

and

$$\mathcal{R}_i = \mathcal{H}_i - D_i - \mathcal{F}_i \ . \tag{B10}$$

B.2 Estimation Algorithm

We proceed to describe the estimation algorithm for the common parameters $\tilde{\boldsymbol{\mu}} = (\omega, \psi, \delta_{\sigma,0}, \delta_{\sigma,1}, \delta_{\eta,0}, \delta_{\eta,1}, \delta_{\xi,0}, \delta_{\xi,1}, \delta_{\zeta,0}, \delta_{\zeta,1})$.

We first define a multi-dimensional grid of parameter values at which we perform a preliminary evaluation of the loss function in problem (11). The grid is defined by the Cartesian product of the following vectors:

$$\begin{aligned} \boldsymbol{\omega}^{(0)} &= [0.1, \quad 1, \quad 10] \\ \boldsymbol{\psi}^{(0)} &= [0.00001, \quad 1, \quad 10] \\ \boldsymbol{\delta}_{\sigma,0}^{(0)} &= [-100, \quad 0] \\ \boldsymbol{\delta}_{\sigma,1}^{(0)} &= [0] \\ \boldsymbol{\delta}_{\eta,0}^{(0)} &= [\ln(0.8) - \ln(0.2), \quad 0, \quad \ln(0.2 - \ln(0.8)] \\ \boldsymbol{\delta}_{\eta,1}^{(0)} &= [0] \\ \boldsymbol{\delta}_{\xi,0}^{(0)} &= [-11.5, \quad \ln(2)] \\ \boldsymbol{\delta}_{\xi,1}^{(0)} &= [0] \\ \boldsymbol{\delta}_{\zeta,0}^{(0)} &= [-11.5, \quad 0] \\ \boldsymbol{\delta}_{\zeta,1}^{(0)} &= [0] \end{aligned}$$

thus including 216 grid-points.

Secondly, we select the grid-points where the value of the loss function is within 0.5% of the identified minimum value and we use them as initial points for the numerical solution to problem (11). We perform a preliminary search using a "Nelder-Mead simplex direct search" method with a high tolerance on the loss function and standard tolerance on parameter values. We then select

the parameters vectors where the value of the loss function is within 1% of the identified minimum value and we use them as initial points for the numerical solution to problem (11) with the same search algorithm and with lower loss function tolerance. We repeat this step twice until we are left with a single candidate optimal parameter vector.

B.3 Objectives and Constraints for Counterfactuals

We first analyze a setting where social planner's objective is to maximize field-specific aggregate scientist's utility, i.e., the sum of field-f scientists' individual utility, by reallocating guaranteed funding G_i and administrative duties D_i within the field, conditional on their total amounts and the total value of additional funding being fixed to the observed endowments.

To this end, we define the output function

$$\mathcal{Y}_i(\tilde{G}_i, \tilde{D}_i, \pi_f) = \alpha_i \left(B_{\min} + G_i + \phi_i \mathcal{F}(\tilde{G}_i, \tilde{D}_i, \pi_f) \right)^{\gamma_i} \mathcal{R}(\tilde{G}_i, \tilde{D}_i, \pi_f)^{1-\gamma_i}$$

where the adjustment factor π_f to fundraising ability in field f enforces the constraint that additional funding in the new, counterfactual allocation must be equal to observed additional funding. For convenience, in this section we omit from the notation the dependence of the policy functions on attributes $\widehat{\boldsymbol{\theta}}_i$ and parameters $\widehat{\boldsymbol{\mu}}_i$.

Therefore, the problem in field f is:

$$\begin{aligned}
\max_{(\tilde{G}_i, \tilde{D}_i)_{i=1}^{N_f}} & \sum_{i=1}^{N_f} \left(\kappa \frac{\mathcal{Y}_i(\tilde{G}_i, \tilde{D}_i, \pi_f)^{1-\eta_i}}{1-\eta_i} - \psi \frac{(\mathcal{H}_i(\tilde{G}_i, \tilde{D}_i, \pi_f) - \tilde{D}_i + \tilde{D}_i^{\xi_i})^{1+\zeta_i}}{1+\zeta_i} \right) \\
& \text{subject to} \\
& \sum_{i=1}^{N_f} \tilde{G}_i = \widehat{G}^{tot} \quad [\text{Multiplier } \lambda_{G,f}] \\
& \sum_{i=1}^{N_f} \tilde{D}_i = \widehat{D}^{tot} \quad [\text{Multiplier } \lambda_{D,f}] \\
& \sum_{i=1}^{N_f} EG_i = \pi_f \sum_{i=1}^{N_f} \widehat{\phi}_i \mathcal{F}_i(\tilde{G}_i, \tilde{D}_i, \pi_f) \ ,
\end{aligned} \tag{B11}$$

where N_f is the number of scientists in field f. We also impose non-negativity constraints on individual $\tilde{G}_i$ and $\tilde{D}_i$ allocations. The individual research output function $\mathcal{Y}_i = \widehat{\alpha}_i \mathcal{B}_i^{\widehat{\gamma}_i} \mathcal{R}_i^{1-\widehat{\gamma}_i}$ and hours $\mathcal{H}_i$ make explicit that both are functions of the considered instruments $(\tilde{G}_i, \tilde{D}_i)$. The last constraint requires that total additional funding in the observed allocations, i.e., the term on the left hand side, equals additional funding in the optimal allocations conditional on the field-specific adjustment factor π_f. Therefore, all the policy functions—i.e., researchers' behavioral responses to the planner's decisions—not only vary with policy levers $(\tilde{G}_i, \tilde{D}_i)$ but also with π_f, which can be interpreted as an endogenous adjustment to fundraising probability. When the new allocation violates the constraint, fundraising probability uniformly shrinks, thus reducing additional funding both directly and indirectly through the behavioral decline in fundraising hours.

In the optimal allocations, the planner seeks to equate the marginal utility of guaranteed funding and duties across individuals, conditional on the non-negativity constraint not being hit and the

additional funding constraint being met. Therefore, the solution of the problem solves, for each $i = 1, ..., N_f$ the following equations:

$$\frac{d\mathcal{V}_i}{d\tilde{G}_i} = \kappa \mathcal{Y}_i^{-\eta_i} \frac{d\mathcal{Y}_i}{d\tilde{G}_i} - \psi(\mathcal{H}_i - \tilde{D}_i + \tilde{D}_i^{\xi_i})^{\zeta_i} \frac{\partial \mathcal{H}_i}{\partial \tilde{G}_i} = \lambda_{G,f} - \lambda_{G^*,i} \tag{B12}$$

$$\frac{d\mathcal{V}_i}{d\tilde{D}_i} = \kappa \mathcal{Y}_i^{-\eta_i} \frac{d\mathcal{Y}_i}{d\tilde{D}_i} - \psi(\mathcal{H}_i - \tilde{D}_i + D_i^{\xi_i})^{\zeta_i} \left(\frac{\partial \mathcal{H}_i}{\partial \tilde{D}_i} - 1 + \xi_i \tilde{D}_i^{\xi_i - 1} \right) = \lambda_{D,f} - \lambda_{D^*,i} \, , \tag{B13}$$

where the marginal product of guaranteed funds $\frac{d\mathcal{Y}_i}{d\tilde{G}_i}$ and the (negative) marginal product of duties $\frac{d\mathcal{Y}_i}{d\tilde{D}_i}$ are determined by equations (B14) and (B15), respectively:

$$\frac{d\mathcal{Y}_i}{dG_i} = \widehat{\alpha}_i \mathcal{B}_i^{\widehat{\gamma}_i} \mathcal{R}_i^{1-\widehat{\gamma}_i} \left(\widehat{\gamma}_i \mathcal{B}_i^{-1} \frac{\partial \mathcal{B}_i}{\partial \tilde{G}_i} + (1 - \widehat{\gamma}_i) \mathcal{R}_i^{-1} \frac{\partial \mathcal{R}_i}{\partial \tilde{G}_i} \right) \tag{B14}$$

$$\frac{d\mathcal{Y}_i}{dD_i} = \widehat{\alpha}_i \mathcal{B}_i^{\widehat{\gamma}_i} \mathcal{R}_i^{1-\widehat{\gamma}_i} \left(\widehat{\gamma}_i \mathcal{B}_i^{-1} \frac{\partial \mathcal{B}_i}{\partial \tilde{D}_i} + (1 - \widehat{\gamma}_i) \mathcal{R}_i^{-1} \frac{\partial \mathcal{R}_i}{\partial \tilde{D}_i} \right) . \tag{B15}$$

Total funding as a function of state variables is defined by $\mathcal{B}_i(G_i, D_i, \pi_f) = B_{\min} + G_i + \pi_f \widehat{\phi}_i \mathcal{F}_i(G_i, D_i, \pi_f)$.

C Appendix: Additional Productivity and Efficiency Results

Table C1: Utility Function Parameter Estimates

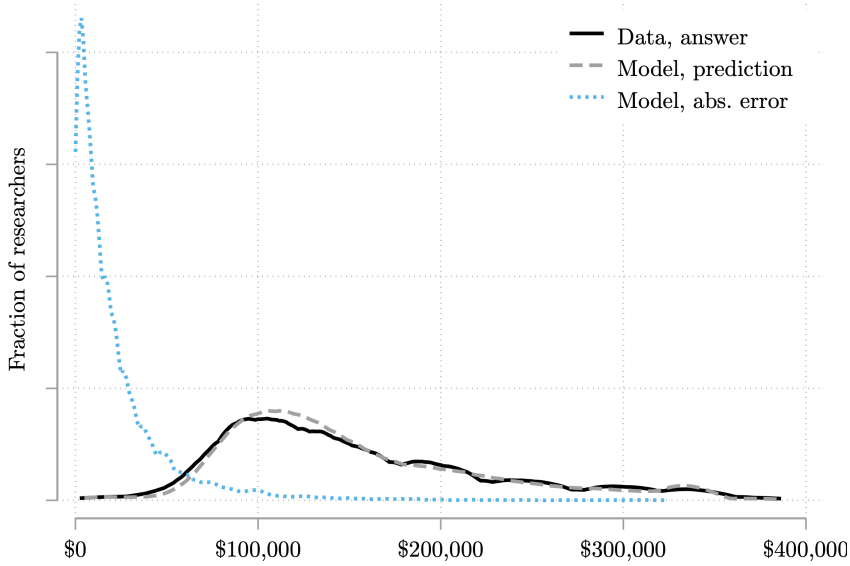
	mean	*s.d.*	$p10$	$p50$	$p90$
Homogenous parameters					
ω	2.7e-04				
σ	1.708				
η	7.2e-04				
ψ	1.0e-14				
ξ	1.503				
ζ	0				
Heterogeneous parameters					
ω	4.2e-04				
σ_i	1.835	0.0338	1.795	1.831	1.882
η_i	0.257	0.0303	0.2217	0.2529	0.2994
ψ	5.9e-15				
ξ_i	1.428	0.0484	1.36	1.432	1.487
ζ_i	3.9e-06	4.0e-07	3.4e-06	3.9e-06	4.4e-06

Note: Reports summary statistics for the parameter estimates depending on whether the model includes homogeneous or heterogeneous parameters; if only means are reported, the parameter is assumed to be homogeneous. Parameter definitions are as follows:

parameter	component
ω	utility from income
σ	utility from income, curvature
η	utility from output, curvature
ψ	disutility from total work time
ξ	disutility from admin., teaching, and other, add.'l curvature
ζ	disutility from total work time, curvature

Figure C1: Model Fit

(a) Answer, Prediction, and Error Distributions

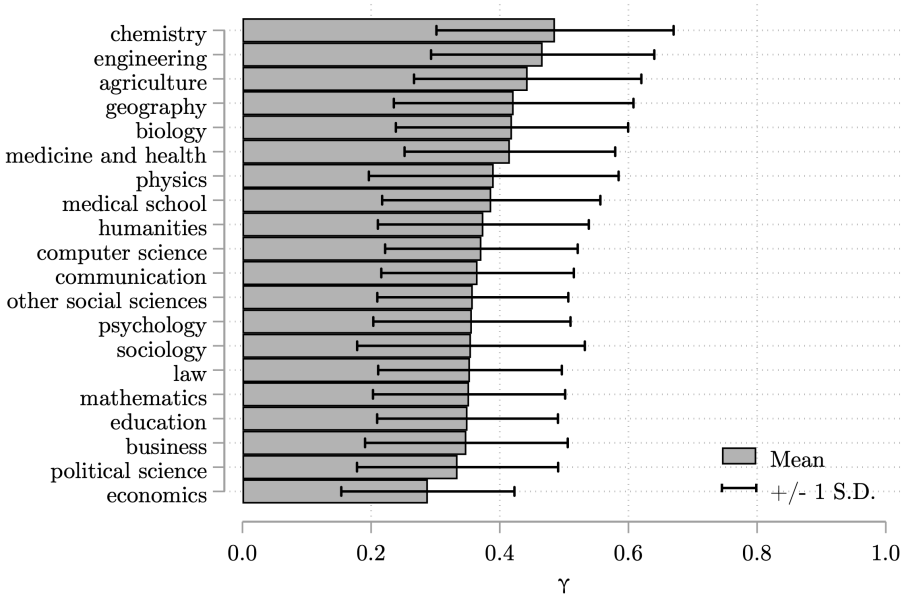


(b) Distributions of Question-specific Errors

	$p10$	$p25$	$p50$	$p75$	$p90$
G +$250,000	0.11	0.88	6.18	15.54	29.58
G +$1,000,000	0.63	3.42	12.25	27.07	53.71
A +20 hr./month	2.19	5.56	11.40	21.01	35.66
A –all hr./month	0.26	2.41	6.17	12.75	22.01

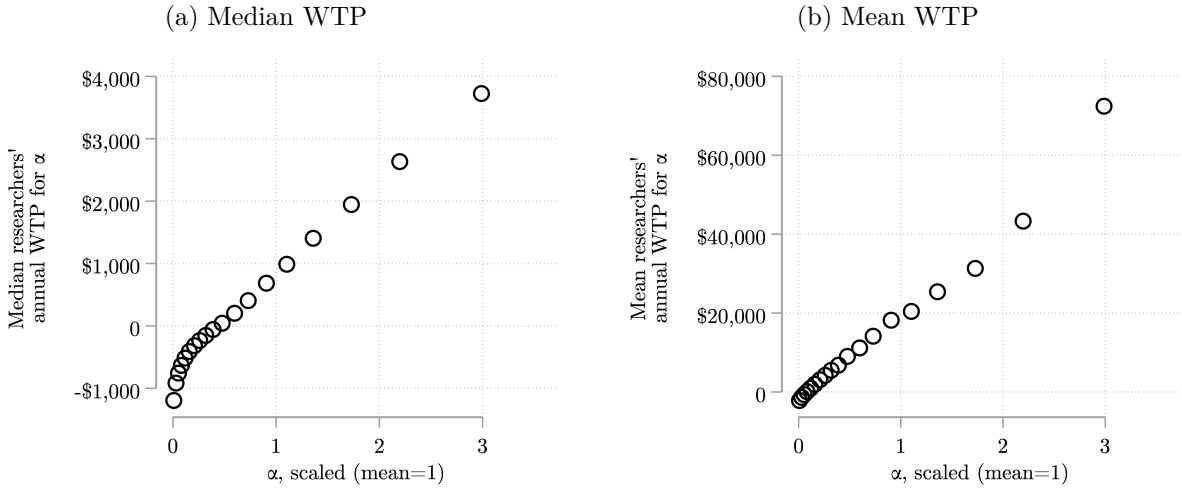
Note: Panel (a) shows the distribution of researchers' responses to the four experiments (i.e., their stated annual salary at indifference) alongside the distribution of the model's predictions and absolute errors. Panel (b) reports summary statistics for the distribution of the absolute errors for each of the four experiments where the model's errors are expressed in percentage points of the actual responses.

Figure C2: Research Funding Intensity by Minor Fields of Study



Note: Shows the mean and standard deviations of researcher-specific estimates of the production function parameter γ_i (funding intensity) split by minor field of study.

Figure C3: Dollar Conversions of Research Productivity



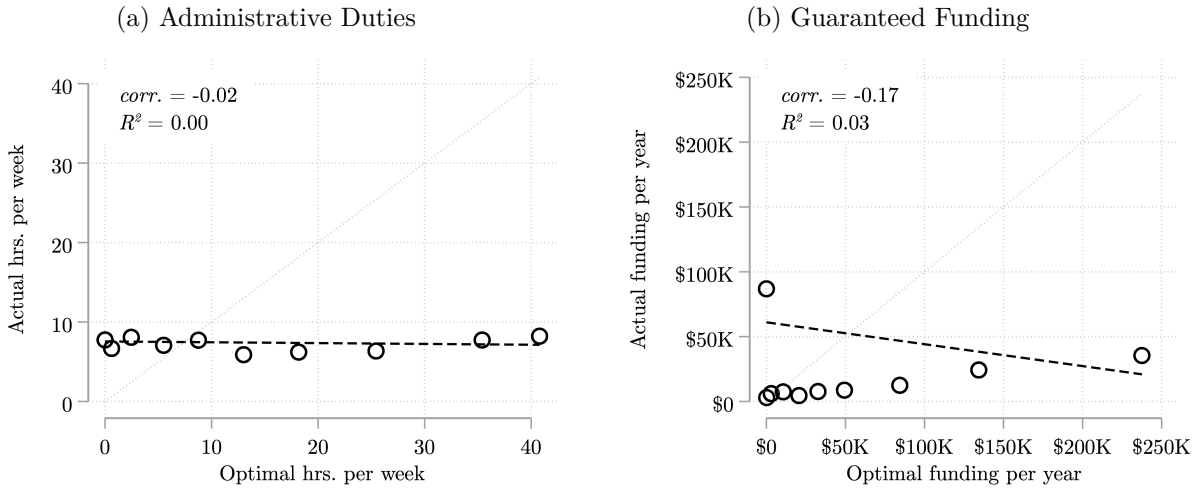
Note: Shows the implied dollar value of productivity in terms of the median (Panel a) or mean (Panel b) of in-sample researchers' willingness to pay for a given productivity level, where productivity (x axis) is scaled so that the sample mean equals one. For example, Panel (a) indicates that the median amount that researchers would be willing to pay to have average productivity is approximately $1,000 per year. Note, negative values imply that researchers must be paid to receive a given productivity level (because it is lower than their current productivity).

Table C2: Observable Output Correlations

	Recent publications, count		Recent publications, cite-weighted	
	(1)	(2)	(3)	(4)
log(α)	0.201***	0.176***	0.276***	0.254***
	(0.0406)	(0.0397)	(0.0648)	(0.0648)
log(recent grant \$)		0.0220***		0.0174***
		(0.00350)		(0.00525)
Field–FE, $\gamma, \phi, \mathbf{X}$–index	✓	✓	✓	✓
R^2	0.18	0.20	0.15	0.16
N obs.	2,703	2,703	2,703	2,703

Note: Reports the estimates from regressions of researchers' publication measures (including publications from 2018–2022) on their estimated research productivity (α) as well as vector of controls that includes major field fixed effects (*Field-FE*) and the researchers' funding intensity (γ), fundraising efficiency (ϕ), their type ($\mathbf{X}$-*index*), and, in some specifications, a control for their publicly observable research grant funding over the same period. Robust standard errors reported; $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Figure C4: Actual Versus Optimal Input Correlations—Policy Levers



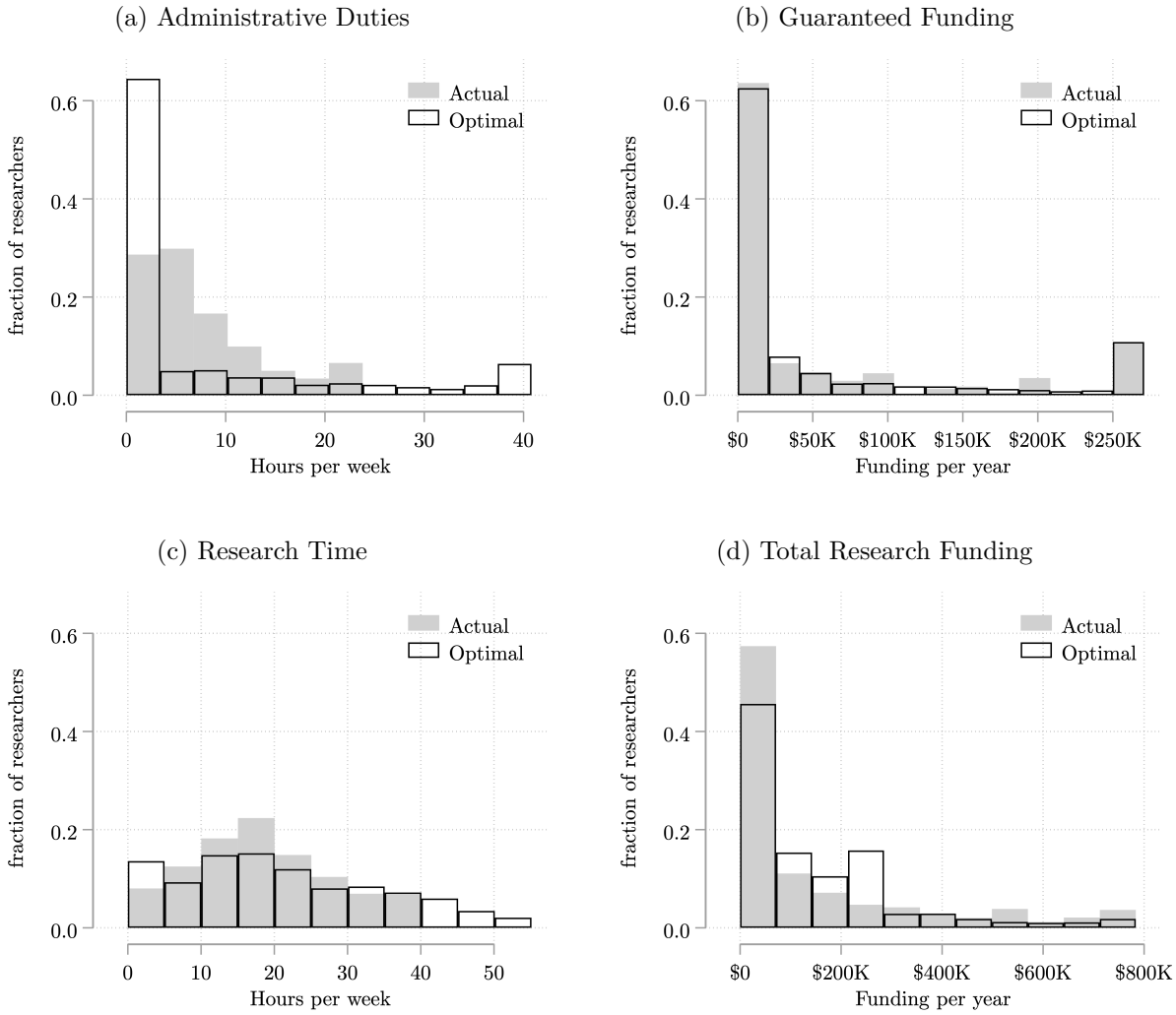
Note: Shows the binned scatterplot of actual and optimal input levels also reporting the pairwise correlation between the two variables and the R^2 statistic from a regression of the actual level on the optimal level. Based on the counterfactual described in Column (4) of Table 3.

Table C3: Production with Actual and Optimized Allocations—Alternative Specifications

	Current allocation	Optimized allocations					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Research inputs							
Research hrs./week, avg.	18.5	+14%	+30%	+6.5%	+14%	+27%	+41%
Research hrs./week, s.d.	9.6	+43%	+60%	+27%	+44%	+43%	+40%
Budget $-K/year, avg.	147.1	0%	0%	0%	0%	0%	+0.0%
Budget $-K/year, s.d.	206.9	–20%	–22%	–22%	–24%	–18%	–21%
Research output							
Output, avg.	*n.r.*	+160%	+160%	+250%	+170%	+160%	+160%
Output, s.d.	*n.r.*	+44%	+45%	+180%	+44%	+44%	+44%
Output per hr.	*n.r.*	+130%	+100%	+230%	+130%	+100%	+81%
Output per $	*n.r.*	+160%	+160%	+250%	+170%	+160%	+160%
Value							
Researcher utility, avg.	*n.r.*	+4.7%	+4.8%	+250%	+4.8%	+4.5%	+2.0%
Researcher utility, s.d.	*n.r.*	–5.4%	–5.1%	+180%	–5.5%	–5.2%	–2.1%
Researcher utility per hr.	*n.r.*	+16%	+27%	+230%	+17%	+25%	+31%
Researcher utility per $	*n.r.*	+4.8%	+4.8%	+250%	+4.6%	+4.5%	+2.0%
Social value, avg.	*n.r.*	+16%	+17%	+250%	+17%	+16%	+6,900%
Social value, s.d.	*n.r.*	+20%	+20%	+180%	+20%	+21%	+60%
Social value per hr.	*n.r.*	+26%	+36%	+230%	+27%	+34%	+4,850%
Social value per $	*n.r.*	+16%	+17%	+250%	+17%	+16%	+6,900%
Input reallocation							
Research hrs./week		47%	37%	46%	44%	43%	42%
Budget $-K/year		44%	44%	38%	38%	50%	53%
Objective, max		U	U	U	U	W	W
Reallocate D		✓	✓	✓	✓	✓	✓
Reallocate G		✓	✓	✓	✓	✓	✓
Absolute input cap			✓				
Homog. param.				✓			
Across-field B					✓		
κ		10	10	10	10	10	100

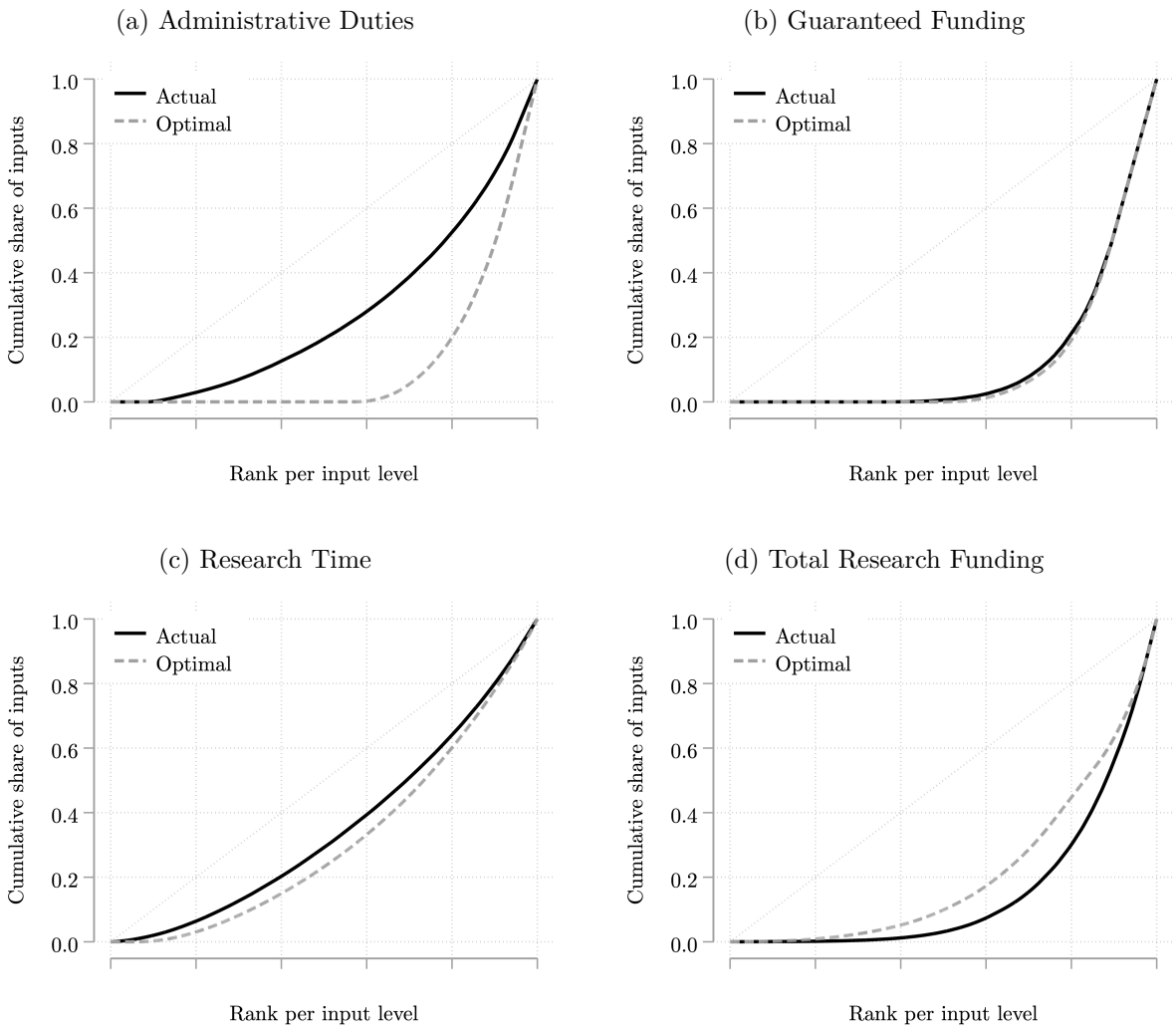
Note: Reports summary statistics for inputs under current allocations (Col. 1). The first three sets of rows in Columns 2–7 report the percentage change in research inputs (*Research inputs*), outputs (*Research outputs*), and utility (*Value*) under alternative allocations. The *Input Reallocation* rows report the amount of inputs reallocated expressed as a percentage of the total level of the input (e.g., 50% implies that half of all dollars are moved from one researcher to another). The bottom sets of rows outline the objective and constraints of the six different counterfactual allocations explored in Columns 2–7. The three different objectives explored are maximizing researchers' private utility (U), social welfare (W), or output (Y). D refers to administrative duties, and G refers to guaranteed research funding. *Unconstrained B* indicates the scenario when the total research budget is left unconstrained and so the total amount of funding in the market is limited only by researchers' fundraising choices. All optimized allocations allow for researchers' behavioral responses after D and/or G have been reallocated. κ indicates the assumed externality multiplier on researchers' production. *Absolute input cap* refers to the scenario where researchers' behavioral responses are limited with a global absolute maximum (as opposed to a relative maximum in the preferred scenario). *Homogeneous param.* refers to the scenario where all non-production-function parameters are homogeneous. *Across field B* allows for the reallocation of inputs across the five major fields of study (which is not allowed in the preferred scenario).

Figure C5: Actual Versus Optimal Input Levels



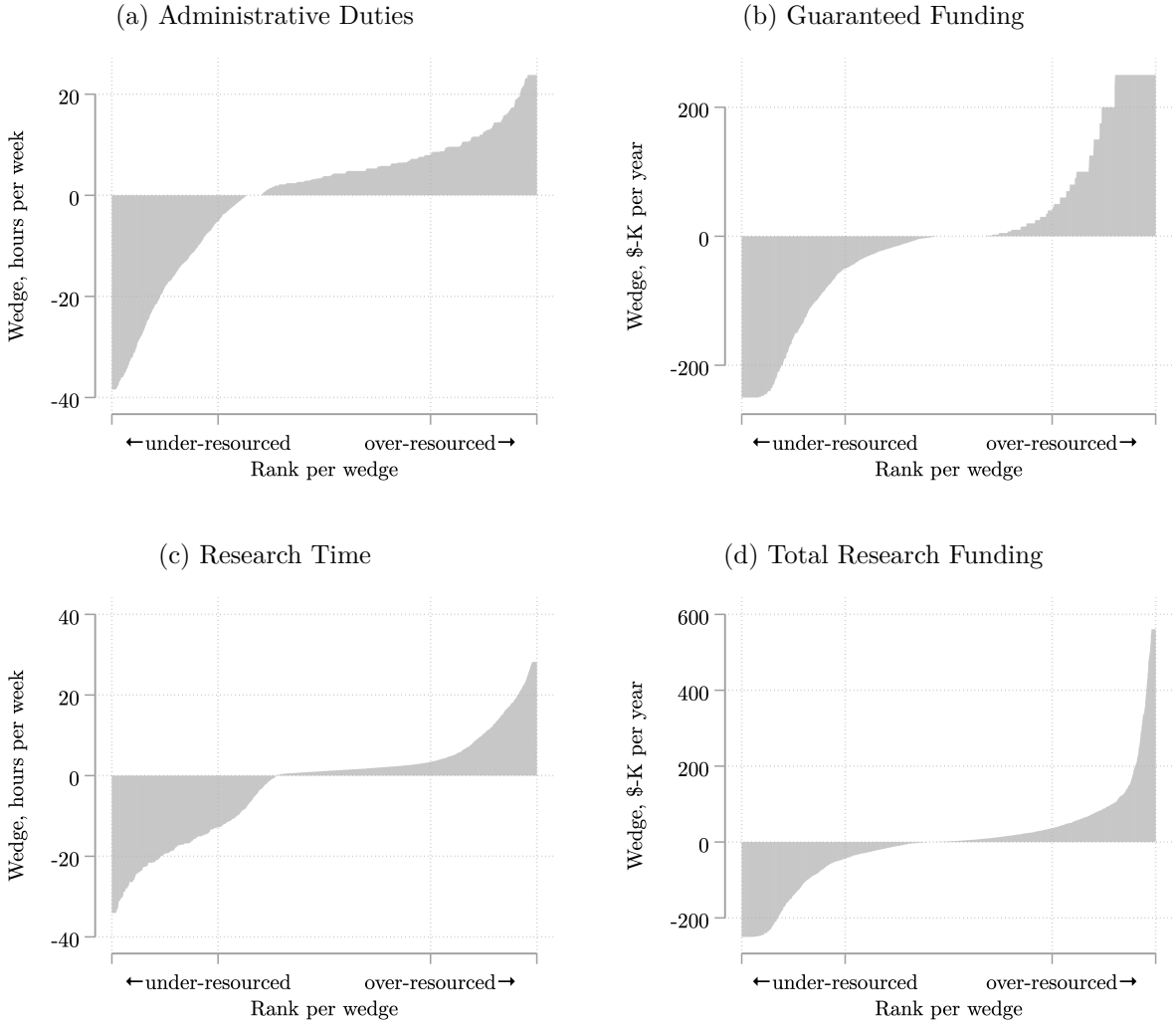
Note: Shows the distribution of actual and optimal input levels. Panels (a–b) show the distribution of the inputs assumed to be directly under the planners' control. Panels (c–d) show the distribution of inputs after researchers' choose their time allocations in response to the planner.

Figure C6: Lorenz Curves for Actual and Optimal Input Levels



Note: Shows Lorenz curves for actual and optimal input levels. Researchers are sorted on the x axis per their ranking in terms of how much of each input they have, and the lines plot the cumulative share of total inputs (per the y axis) summing from the lowest- to highest-ranked researcher. Thus, plots closer to the 45^o line indicate input allocations closer to equality.

Figure C7: Input Wedge Distributions



Note: Shows the distribution of the input wedge, which equals a researcher's actual input level minus their optimal input level; positive values indicate the researcher is over-resourced and negative values indicate the researcher is under-resourced relative to the optimum.

Table C4: Input Wedges per Common Productivity Proxies

	(1)	(2)	(3)	(4)	(5)
Panel (a): Actual research time					
Optimal level	0.398***	0.400***	0.396***	0.402***	0.403***
	(0.0169)	(0.0168)	(0.0169)	(0.0164)	(0.0164)
Recent research funding, own		0.138***			
		(0.0345)			
Recent research funding, institution			–0.00220		
			(0.0213)		
Recent publications				0.209***	
				(0.0191)	
Recent citations					0.178***
					(0.0231)
R^2	0.16	0.18	0.16	0.20	0.19
N obs.	3,072	3,072	3,057	3,072	3,072
Panel (b): Actual research funding					
Optimal level	0.755***	0.731***	0.755***	0.718***	0.734***
	(0.00895)	(0.0117)	(0.00896)	(0.0114)	(0.0103)
Recent research funding, own		0.111***			
		(0.0369)			
Recent research funding, institution			0.0105		
			(0.0168)		
Recent publications				0.129***	
				(0.0186)	
Recent citations					0.105***
					(0.0189)
R^2	0.57	0.58	0.57	0.59	0.58
N obs.	3,072	3,072	3,057	3,072	3,072

Note: Reports results from regressions of actual input levels on optimal input levels and common proxies for researchers' producitivities. All variables are standardized. All proxies are based on data from one to two years prior to the survey. Robust standard errors reported; $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.